

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

ESCUELA ACADÉMICO PROFESIONAL DE ESTADÍSTICA

Modelo de clases latentes multinivel aplicado a un caso

TESIS

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Melissa Sheedy Cotrina Salas

Lima – Perú

2014

MODELO DE CLASES LATENTES MULTINIVEL

APLICADO A UN CASO

MELISSA SHEEDY COTRINA SALAS

Tesis presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciada en Estadística.

Aprobada por:

Mg. Ana María Cárdenas Rojas
(Presidente de jurado)

Mg. Rosa Ysabel Adriazola Cruz
(Jurado)

Mg. Emma Norma Cambillo Moyano
(Jurado Miembro asesor)

Lima - Perú
2014

FICHA CATALOGRÁFICA

COTRINA SALAS, MELISSA SHEEDY

**Modelo de Clases Latentes Multinivel
aplicado a un caso.**

ix, 115p. , 29,7 cm. (UNMSM, Licenciada,
Estadística, 2014).

Tesis, Universidad Nacional Mayor de San
Marcos, Facultad de Ciencias Matemáticas 1.

Estadística I. UNMSM/F de CM II. Modelo de
Clases Latentes Multinivel aplicado a un caso

DEDICATORIA

A mis padres
quiénes me brindaron
su constante apoyo e
incondicional amor.

AGRADECIMIENTO

A Dios, por su inmenso amor; quién es mi luz y mi fuerza en tiempos difíciles.

A mi asesora Emma Cambillo Moyano, por su paciencia y arduo trabajo mostrado durante el desarrollo de este trabajo.

Al profesor Jeroen K. Vermunt, por facilitarme el software Latent GOLD 5.0, utilizado en la aplicación de la investigación.

A mis padres y familiares por todo el esfuerzo realizado.

RESUMEN

Modelo de Clases Latentes Multinivel aplicado a un caso

MELISSA SHEEDY COTRINA SALAS

SEPTIEMBRE 2014

Orientador: Mg. Emma Cambillo Moyano

Título obtenido: Licenciada en Estadística

El presente trabajo tiene como objetivo investigar el Modelo de Clases Latentes Multinivel (MCLM). Este modelo es una extensión muy importante del Modelo de Clases Latentes (MCL), cuyo objetivo principal es el de construir perfiles, a partir de un conjunto de variables categóricas observadas; bajo el supuesto de independencia entre observaciones, lo que frecuentemente es vulnerado si se analizan datos que presentan cierta jerarquía. El MCLM es un modelo que permite analizar los datos modificando este supuesto, al incorporar variables latentes discretas en todos los niveles de la jerarquía, con la finalidad de conseguir una solución óptima del número de subclases o perfiles que pertenecen a cada una de estas variables latentes.

Finalmente se utilizó el MCLM para determinar los niveles de depresión de una muestra de 399 internos pertenecientes a ocho establecimientos penitenciarios del país, Para el análisis se consideró un MCLM con tres clases latentes en el primer nivel, tomando en cuenta los resultados anteriores se consideró segmentar a los establecimientos penitenciarios en tres grupos: el primero, conformado por una mayor proporción de internos que fueron clasificados como “moderadamente afectados” por la depresión (58%); el segundo grupo presentó una mayor proporción de internos “altamente afectados” por esta enfermedad (69%) y el último grupo, conformado exclusivamente por internas que en su mayoría no presentaron episodios depresivos (85%).

Palabras-clave: Modelo de Clases Latentes, Algoritmo EM.

ABSTRACT

Multilevel Latent Class Modeling applied to a case

MELISSA SHEEDY COTRINA SALAS

SEPTEMBER 2014

Assessor: Mg. Emma CambilloMoyano

Title obtained: Licensed in Statistics

The goal of this work is to investigate the Modelo Class Multilevel latent (MCML). This model is an important extension Latent Class Model (LCM), whose main objective is to build profiles from a set of categorical variables observed; under the supposed of independence between observations, what is often violated if data analyzes presented certain hierarchy. MCML is a model to analyze data changing this supposed , incorporating discrete latent variables at all levels of the hierarchy, in order to get optimal solution of the number of subclasses or profiles belonging to each of the latent variables.

Finally MCML was used to determine levels of depression a sample of 399 inmates from eight local prisons of the country, for the analysis was considered an MCLM with three latent classes in the first level, taking into account the above results are considered to segment the locals prisons into three groups: the first, consisting of more proportion of inmates who were classified as "moderately affected "by the depression (58%); the second group had a higher proportion of inmates "highly affected" by this disease (69%) and the last group, consisting entirely of inmates who in their Most no showed depressive episodes (85%).

Palabras-clave: ,LatentClassModels, EM.Algorithm

Índice General

CAPÍTULO 1	7
Introducción.....	7
1.1. Objetivos	8
1.2. Organización del Trabajo	9
CAPÍTULO 2	10
Marco teórico	10
2.1. Métodos de estimación	10
2.2. Algoritmo EM.....	21
2.1. Distribución Logística	26
CAPÍTULO 3	28
Modelo de Clases Latentes.....	28
3.1. Introducción.....	28
3.2. Planteamiento del Modelo de Clases Latentes.....	31
3.3. Estimación del Modelo de Clases Latentes	37
3.4. Bondad de ajuste del Modelo de Clases Latentes.....	39
3.5. Criterios de selección del Modelo de Clases Latentes	42
3.6. Modelo de Clases Latentes Multi-grupos.....	43
CAPÍTULO 4	46
Modelo de Clases Latentes Multinivel (MCLM).....	46
4.1. Introducción.....	46
4.2. Definición del modelo	47
4.3. Estimación de Máxima Verosimilitud (EMV)	56
4.4. Evaluación del Modelo de Clases Latentes Multinivel	59
4.5. Selección del Modelo de Clases Latentes Multinivel	60
CAPÍTULO 5	63
Aplicación del Modelo de Clases Latentes Multinivel para la construcción de perfiles de reos pertenecientes a ocho establecimientos penitenciarios del país que sufren episodios depresivos	63
5.1. Definición y descripción de las variables.....	64
5.2. Análisis descriptivo preliminar.....	65
5.3. Modelo de clases latentes para los internos de ocho establecimientos penitenciarios del Perú	67
5.4. Modelo de Clases Latentes Multinivel para examinar la variación a través de los establecimientos penitenciarios	71
CONCLUSIONES Y RECOMENDACIONES.....	80
REFERENCIAS BIBLIOGRÁFICAS	82

APÉNDICE A.....	87
Modelo de Regresión Multinivel	87
1. Introducción.....	88
2. Modelo de Regresión Logística Multinivel	91
APÉNDICE B.....	95
Aplicación del Latent GOLD 5.0 para estimar el Modelo de Clases Latentes Multinivel	95
ANEXO	105

Lista de Abreviaturas

ACL	Análisis de Clases Latentes.
MCL	Modelos de Clases Latentes
MCLM	Modelo de clases latentes Multinivel
EM	Esperanza y maximización
MV	Máxima Verosimilitud
NR	Newton Raphson
AIC	Criterio de Información Akaike
BIC	Criterio de Información Bayesiano
FDP	Función de densidad de probabilidad
CIE	Clasificación Internacional de Enfermedades
LL	Log-Verosimilitud

Índice de figuras

2.1. Objetos y siluetas.....	15
2.2. Ajuste de funciones de distribución normal a histogramas.....	17
2.3. Diferentes formas de la función de Kernel.....	18
2.4. Modelo Mixto Gausiano.....	19
3.1. Matriz de datos.....	31
3.2. Modelo de Clases Latentes con una variable latente y K variables manifiestas.....	35
4.1. Modelo de clases latentes con un efecto aleatorio no paramétrico en la distribución de clases latentes.....	55
4.2. Proceso de modelización del MCLM.....	62
5.1. Modelos de Clases Latentes conteniendo de uno a cuatro clases latentes.....	68
5.2. Gráfico de probabilidades condicionales de respuesta “Sí” a cada ítem, dada la clase latente para el modelo no restringido de tres clases.....	69
5.3. Gráfico de probabilidades condicionales de respuesta “Sí” a cada ítem, dada la clase latente para el modelo no restringido de cuatro clases.....	70
5.4. Modelo de clases latentes multinivel con tres clases en el nivel 1 y tres clases en el nivel 2.....	73
5.5. Distribución de encuestados entre los clusters latentes dentro de cada clase del segundo nivel.....	76
B.1. Archivo de datos en SPSS aplicación.sav.....	95
B.2. Cuadro de diálogos Open File.....	96
B.3. Ventana principal del programa Latent GOLD 5.0.....	96
B.4. Menú de selección del modelo.....	97
B.5. Cuadro de diálogos para analizar el Modelo de Clases Latentes para el primer nivel.....	98
B.6. Ingreso de variables	98
B.7. Cuadro de diálogos previo a la estimación del modelo.....	99
B.8. Resumen del modelo.....	100
B.9. Estimación de parámetros del MCL con tres clases	101
B.10 Perfil obtenido en el MCL con tres clases.....	101
B.11. Gráfico de perfiles para el MCL con tres clases.....	102

B.12. Cuadro de diálogos para la estimación del MCLM.....	103
B.13. Cuadro de diálogos previo a la estimación del MCLM de 1 a 4 clases en el segundo nivel.....	104

Índice de cuadros

1.1.	Clasificación de modelos con variables latentes.....	7
2.1.	Probabilidades a priori P(Objeto).....	15
2.2.	Probabilidades condicionales P(Silueta/Objeto).....	15
3.1.	Clasificación de modelos con variables latentes	28
3.2.	Estadísticos de bondad de ajuste.....	41
5.1.	Variables relacionados a episodios depresivos según el CIE-10.....	64
5.2.	Variables relacionadas a características de los internos.....	67
5.3.	Estadísticos descriptivos de la edad de los internos.....	67
5.4.	Estadísticos acerca de las características de los internos.....	68
5.5.	Estadísticos de ajuste para modelos conteniendo de uno a cuatro de clases latentes.....	69
5.6.	Estadísticos descriptivos según establecimiento penitenciario.....	72
5.7.	Ajuste de modelos para la variación entre los centros penitenciarios (unidades del segundo nivel).....	72
5.8.	Resultados del modelo de tres clases latentes en el nivel 2.....	74
5.9.	Tamaño relativo de los tres clusters (clases latentes) y las probabilidades de respuesta “SÍ” y “NO” para cada ítem, dado la membresía del cluster.....	75
5.10	Tamaño de las clases del segundo nivel y distribución de encuestados entre los clusters dentro de cada clase del segundo nivel.....	76
5.11.	Características demográficas de los internos por grupos formados en el segundo nivel.....	79

CAPÍTULO 1

Introducción

En el análisis de tipo cuantitativo es muy frecuente observar el uso de instrumentos de medición, a fin de poder estudiar variables causales importantes que no son directamente observables, incorporando de esta forma constructos latentes tales como actitudes y percepciones, donde los encuestados responden a un conjunto de preguntas las cuales son manifestaciones de la variable latente implícita. Por tal motivo se puede definir a la variable latente como una variable que no se incluye entre las variables estudiadas; sin embargo, tiene un importante efecto sobre la relación que existe entre ellas (Moore, 2005).

En la actualidad existen diferentes tipos de modelos que incorporan variables latentes. Según Bartholomew y Knoop (1999) la selección del modelo dependerá de la escala de medición tanto de las variables observadas como el de las variables latentes. En el Cuadro N°1.1 se presenta la clasificación de estos modelos.

Cuadro 1.1. Clasificación de modelos con variables latentes

		Variables manifiestas	
		Continuas	Categóricas
Variables latentes	Continuas	Análisis factorial	Análisis de rasgo latente
	Categóricas	Análisis de perfil latente	Análisis de clases latentes

El Análisis de Clases Latentes (ACL) permite identificar un conjunto de clases latentes exhaustivas y mutuamente excluyentes a partir de un conjunto de

variables observadas categóricas (Goodman, 1974); una clase latente es un subgrupo no observable, el cual se construye a partir de casos relacionados entre las variables observadas, las que se utilizan para estimar los parámetros del modelo; por lo que el Modelo de Clases Latentes (MCL) puede ser visto como una método de clasificación probabilística.

Al analizar conjuntos de datos obtenidos a través de métodos de muestreo para encuestas complejas, donde la probabilidad de selección no es la misma para todos los miembros de la población, surge un problema en la aplicación de los MCL, cuando se incumple el supuesto de independencia entre las unidades de análisis.

1.1. Objetivos

El objetivo general de esta tesis es investigar el Modelo de Clases Latentes Multinivel considerando su aplicación a un conjunto de datos reales. De manera específica:

- ✓ Revisar la literatura acerca del Modelo de Clases Latentes (MCL) y del Modelo de Clases Latentes Multinivel (MCLM).
- ✓ Presentar el MCLM.
- ✓ Investigar los diferentes métodos de estimación del MCL y MCLM.
- ✓ Evaluar el MCL y MCLM.
- ✓ Aplicar el modelo a un conjunto de datos reales.

1.2. Organización del Trabajo

El Capítulo 2, está centrado en la presentación de un conjunto de definiciones que serán utilizadas en los siguientes capítulos de la presente investigación, una revisión de los métodos de estimación, el algoritmo EM, y la distribución logística.

En el Capítulo 3, se muestra el desarrollo teórico del MCL en base a una revisión exhaustiva de la literatura correspondiente, a fin de explicar los conceptos necesarios que permitirán una mayor comprensión del modelo propuesto, el cual hace posible la modificación de uno de los supuestos del MCL.

El Capítulo 4 se centra en la presentación de los fundamentos teóricos del MCLM, mostrando la estimación, evaluación y selección del modelo.

En el Capítulo 5 se realiza una aplicación del MCLM a un conjunto de datos reales.

Finalmente se presentan las principales conclusiones y resultados del trabajo; así mismo, se sugiere futuras investigaciones que pueden derivar de este trabajo.

CAPÍTULO 2

Marco teórico

En este capítulo se presentan las definiciones básicas que se emplearán en el desarrollo de la tesis, tales como “métodos de estimación”, “algoritmo EM”, “distribución logística”; las cuales serán utilizadas en los siguientes capítulos para una mejor comprensión del modelo.

2.1. Métodos de estimación

La estadística provee métodos que permiten obtener estimadores que mejor representen a los verdaderos valores de los parámetros a partir de un subconjunto de la población (muestra); así mismo, cuantifica el error asociado a la estimación.

La mayoría de distribuciones de probabilidad dependen de cierto número de parámetros, aquellos que son estimados a partir de los datos cuya distribución es conocida.

Estimación Paramétrica

Bajo el supuesto que los datos provienen de una distribución específica (por ejemplo, para variables aleatorias continuas se asume una distribución normal) en la población implícita; por lo que, los parámetros de localización y escala del modelo son entonces ajustados a dicha distribución. Desafortunadamente, en muchos casos una elección a priori de la función de distribución del modelo no

es el adecuado, proporcionando así, una falsa representación de la verdadera función de distribución de los datos.

La estimación de los parámetros mediante este enfoque se presenta de dos formas:

Estimación máxima verosimilitud

La estimación Máxima Verosimilitud (MV) es un método utilizado para la obtención de estimadores, lo cual supone que la distribución de probabilidad de las observaciones es conocida.

Sea $X = (X_1, X_2, \dots, X_n)$ una muestra de variables aleatorias seleccionadas de acuerdo a una familia de probabilidades P_θ ; así mismo, $f(x; \theta)$ y $x = (x_1, x_2, \dots, x_n)$ serán utilizadas para denotar la función de densidad de n variables independientes e idénticamente distribuidas cuando θ es una constante desconocida y fija, la cual pertenece al conjunto $\Theta \subset R^k$.

Definición 1. La función de verosimilitud es considerada como una función de θ que es definida como:

$$L(\theta; x) = f(x, \theta)$$

x es considerada como fija y θ , como la variable para L . La función *log-verosimilitud* es definida como $l(\theta; x) = \log L(\theta; x)$.

Definición 2. El estimador de *máxima verosimilitud* es el valor $\hat{\theta} = \hat{\theta}(x) \in \Theta$ maximizando $L(\theta; x)$; siempre que,

$$L(\hat{\theta}(x)) = \max_{\theta \in \Theta} L(\theta; x)$$

Los estimadores de máxima verosimilitud poseen propiedades importantes, especialmente cuando el tamaño de muestra es grande.

Esta clase de estimadores posee una propiedad importante, si $\hat{\theta}(x)$ es un estimador de máxima verosimilitud para θ , entonces $g(\hat{\theta}(x))$ es un estimador de máxima verosimilitud para $g(\theta)$. Por ejemplo, si θ es un parámetro de la varianza y $\hat{\theta}$ es el estimador de máxima verosimilitud, entonces $\sqrt{\hat{\theta}}$ es el estimador de máxima verosimilitud de la desviación estándar. Esta flexibilidad en el criterio de estimación sólo es posible para el caso de estimadores insesgados.

Ejemplo

Sea el experimento que consiste de n ensayos Bernoulli con probabilidad de éxito p ,

$$L(p; x) = p^{x_1}(1-p)^{(1-x_1)} \dots p^{x_n}(1-p)^{(1-x_n)} = p^{(x_1+\dots+x_n)}(1-p)^{n-(x_1+\dots+x_n)}$$

$$\ln L(p; x) = \ln p \left(\sum_{i=1}^n x_i \right) + \ln(1-p) \left(n - \sum_{i=1}^n x_i \right) = n(\bar{x} \ln p + (1-\bar{x}) \ln(1-p)).$$

$$\frac{d}{dp} \ln L(p; x) = 0$$

$$n \left(\frac{\bar{x}}{p} - \frac{1-\bar{x}}{1-p} \right) = 0$$

$$n \frac{\bar{x} - p}{p(1-p)} = 0$$

$$\hat{p} = \bar{x}$$

Así, el estimador de máxima verosimilitud $\hat{p} = \bar{x}$.

Inferencia Bayesiana

En contraste con el método de estimación MV, surge la inferencia bayesiana como una nueva alternativa de estimar parámetros, el cual permite combinar diferentes fuentes de información sin tomar en cuenta grandes tamaños de muestra, lográndose un análisis eficaz de problemas con estructuras complejas al proporcionar un entorno más óptimo al realizar comparaciones múltiples; extendiéndose así a modelos más flexibles con nuevas formas de aproximación.

A diferencia de la definición clásica de la probabilidad, en esta forma de estimación se considera también a la misma como el resultado de una construcción mental que el observador presenta acerca de una afirmación, donde dicha certidumbre se ciñe a los axiomas que forman parte de la teoría de probabilidades; en este sentido, las probabilidades varían de un sujeto a otro, por lo cual suelen llamarse *certidumbres personales*, *credibilidad*, *probabilidades personales* o *grados de creencia*.

El teorema de Bayes es el recurso probabilístico que se utiliza para actualizar la asignación probabilística inicial o a priori $P(H)$, de una hipótesis H a una probabilidad actualizada o a posteriori $P(H/D)$, basada en una nueva observación D . Este teorema se origina a partir de la definición de probabilidad condicional. Sean dos sucesos A y B , la probabilidad condicional de A dado B se presenta de la siguiente forma:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (2.2)$$

Análogamente

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad (2.3)$$

Reemplazando $P(A \cap B) = P(B/A)P(A)$ en (2.3), se obtiene la regla de Bayes:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad (2.4)$$

Al obtener k sucesos mutuamente excluyentes A_1, A_2, \dots, A_k , la *ley de la probabilidad total* puede ser expresada como:

$$P(B) = \sum_{i=1}^k P(B/A_i)P(A_i)$$

Por lo tanto, otra forma de presentar la regla de Bayes sería de la siguiente forma:

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{\sum_{i=1}^k P(B/A_i)P(A_i)} \quad (2.5)$$

Así mismo el denominador puede ser considerado como una constante de normalización:

$$P(A/B) = c * P(B/A)P(A)$$

Cuando $k=2$, la ecuación (2.5) se reduce a:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/\bar{A})P(\bar{A})}$$

Ejemplo

Para ilustrar este método de estimación se mostrará el ejemplo realizado por Eshky (2008).

Dada la silueta de un objeto, se necesita inferir que objeto es.

La probabilidad de seleccionar el objeto (distribución a priori), $P(\text{Objeto})=P(A)$, es

Cuadro 2.1. Probabilidades a priori $P(\text{Objeto})$

Objeto	Probabilidad
Cubo	0.3
Cilindro	0.2
Esfera	0.1
Prisma	0.4

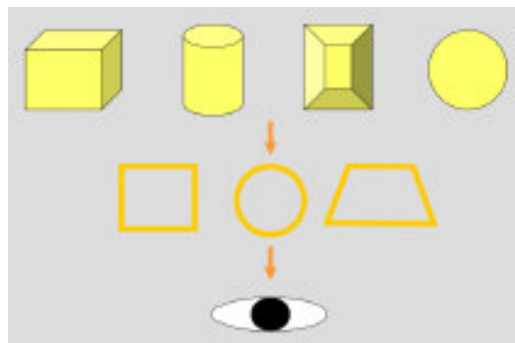


Figura 2.1. **Objetos y siluetas**

La probabilidad de una silueta dado un objeto, $P(\text{Silueta}/\text{Objeto})=P(B/A)$ es presentada en el Cuadro 2.2.

Cuadro 2.2. Probabilidades condicionales $P(\text{Silueta}/\text{Objeto})$

Silueta	Objeto			
	Cubo	Cilindro	Esfera	Prisma
Cuadrado	1.0	0.6	0.0	0.4
Círculo	0.0	0.4	1.0	0.0
Trapezoide	0.0	0.0	0.0	0.6

Calculando el valor de la constante de normalización, c:

$$c = 1 / \left[\sum_{i=1}^k P(B/A_i)P(A_i) \right]$$

$$c = 1 / [(1.0)(0.3) + (0.6)(0.2) + (0.0)(0.1) + (0.4)(0.4)]$$

$$c = 1.72$$

Entonces la constante de normalización c es 1.72.

La distribución posterior de los objetos dada las siluetas $P(\text{objeto/silueta})=P(A/B)$ puede ser calculada.

Si A=Cuadrado

$$P(\text{Cubo/Cuadrado}) = (c)(1.0)(0.3) = 0.52$$

$$P(\text{Cilindro/Cuadrado}) = (c)(0.6)(0.2) = 0.21$$

$$P(\text{Esfera/Cuadrado}) = (c)(0.0)(0.1) = 0.00$$

$$P(\text{Prisma/Cuadrado}) = (c)(0.4)(0.4) = 0.28$$

Estimación No paramétrica

La estimación no paramétrica no requiere información acerca de la forma de la distribución de la variable de estudio, ni de los parámetros de la distribución supuesta. En la estimación no paramétrica, el número de parámetros crece con el tamaño del conjunto de datos; por lo tanto, los modelos pueden llegar rápidamente a ser difíciles de manejar y pueden tomar demasiado tiempo para calcular las estimaciones si se incorpora nuevos datos.

La función de Kernel permite describir una variable categórica o una variable continua considerando valores discretos (como la edad medida en años), para lo cual frecuentemente se emplea un histograma. Esta suavización conecta los puntos medios del histograma

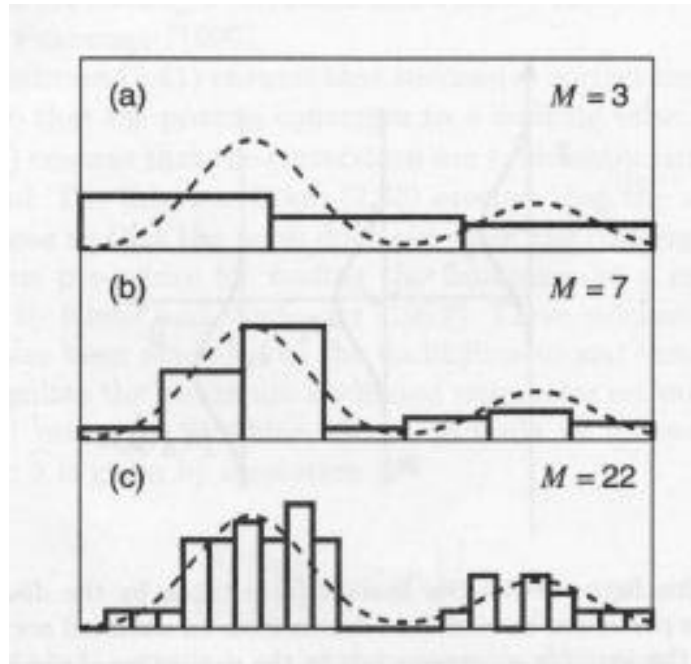


Figura 2.2. **Ajuste de funciones de distribución normal a histogramas**

M: Número de intervalos.

Sea $f(x)$ la función de densidad para una variable continua, el estimador de densidad de Kernel de $f(x)$ en $x = x_0$ (Baum, 2013) está dado por:

$$\hat{f}(x_0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x_0}{h}\right)$$

h representa la amplitud del intervalo y $K(x)$, una función de kernel, definida como una función de densidad de probabilidad (FDP) simétrica alrededor de cero que satisface las siguientes condiciones:

$$\int_{-\infty}^{\infty} K(x)dx = 1; \quad K(x) = K(-x)$$

Según Izenman (1991), el estimador *Parzen window*, selecciona una función de kernel bien definida en cada punto de los datos con una constante optimizada.

En la práctica, las funciones kernel gaussianas son las más utilizadas, estas funciones son estimadas al ser calculadas promediando las densidades gaussianas en cada punto de datos. Una ventaja de utilizar esta técnica es no asumir alguna forma funcional de la FDP, lo cual permite que su forma sea enteramente determinado por los datos.

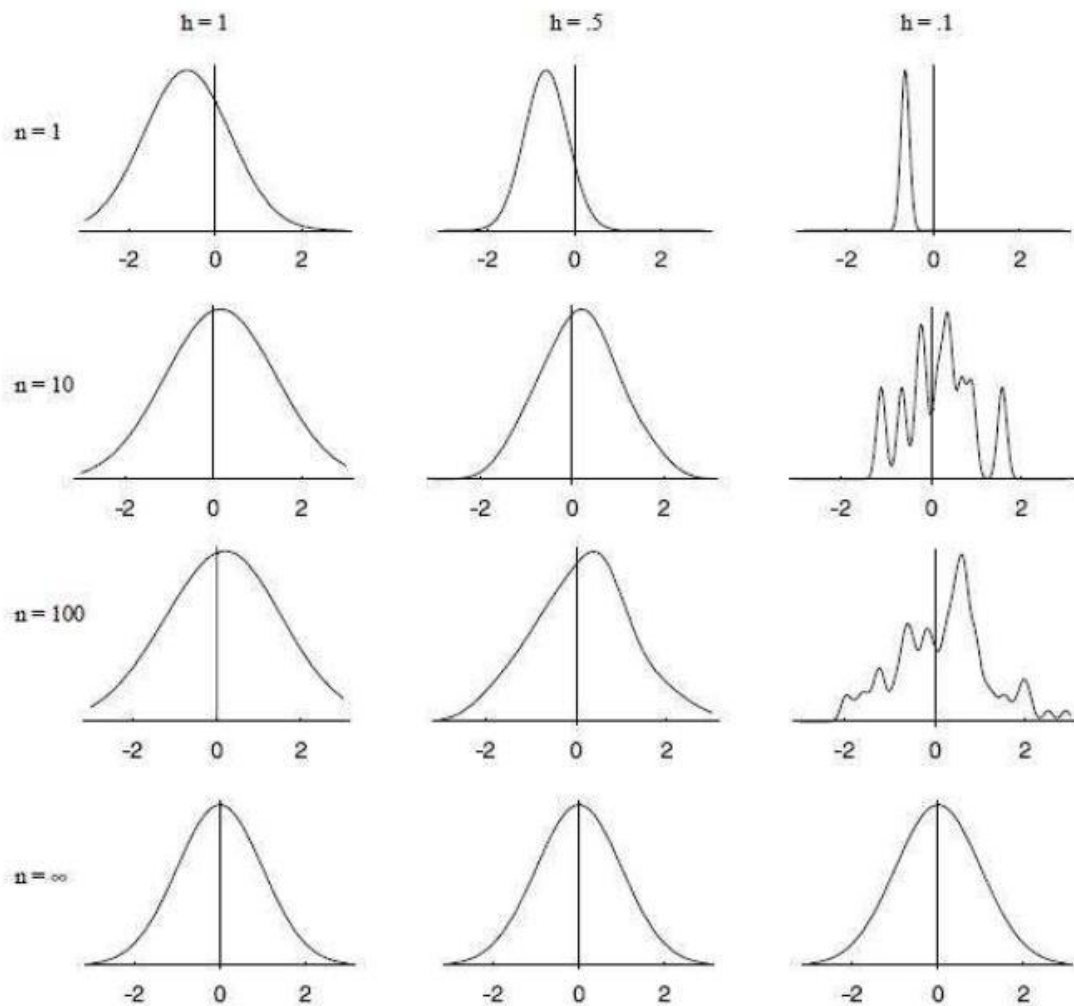


Figura 2.3. Diferentes formas de la función de Kernel

Estimación Semiparamétrica

Este método de estimación es muy utilizado en modelos mixtos a fin de poder alcanzar lo mejor de los métodos ya antes mencionados, permitiendo una forma funcional general para densidades donde el número de parámetros se pueda extender independientemente del tamaño del conjunto de los datos (Vipul, 2010).

Al igual que la estimación no paramétrica, en este método no se asume una forma a priori de la FDP; sin embargo, posee una mayor ventaja, porque evita un crecimiento descontrolado del número de parámetros respecto al tamaño del conjunto de datos; así mismo, este método de estimación permite manejar grandes conjuntos de datos, lo cual es aplicado en problemas de aprendizaje en tiempo real.

Ejemplo

Un claro ejemplo de este enfoque de estimación son los modelos mixtos gaussianos.

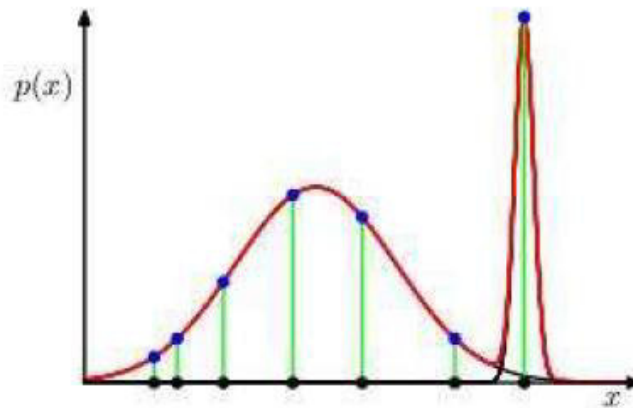


Figura 2.4. **Modelo Mixto Gaussiano**

$$p(x/w, \mu, \Sigma) = \sum_{k=1}^2 w_k N(x/\mu_k, \Sigma_k)$$

$$\sum_{k=1}^K w_k = 1$$

de igual modo, $p(x)$ se puede expresar como:

$$p(x) = \sum_{k=1}^K P(k^\circ \text{mixto}) P(x/k^\circ \text{mixto})$$

$$k^\circ \text{mixto} \sim N(\mu_k, \Sigma_k)$$

Por definición $P(x/k^\circ \text{mixto})$ es igual a:

$$P(x/k^\circ \text{mixto}) = \frac{P(x, k^\circ \text{mixto})}{P(k^\circ \text{mixto})}$$

Entonces $P(x, k^\circ \text{mixto}) = P(x/k^\circ \text{mixto}) * P(k^\circ \text{mixto})$

Así mismo:

$$P(k^\circ \text{mixto}/x) = \frac{P(x, k^\circ \text{mixto})}{P(x)}$$

$$P(k^\circ \text{mixto}/x) = \frac{P(x/k^\circ \text{mixto}) * P(k^\circ \text{mixto})}{\sum_{k=1}^K P(k^\circ \text{mixto}) P(x/k^\circ \text{mixto})}$$

$$P(k^\circ \text{mixto}/x) = \gamma(z_k) = \frac{w_k N(x/\mu_k, \Sigma_k)}{\sum_{k=1}^K w_k N(x/\mu_k, \Sigma_k)}$$

Al maximizar la log verosimilitud de forma iterativa se obtiene lo siguiente

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k) (x_n - \mu_k)^T$$

$$w_k = \frac{N_k}{N}$$

donde $N_k = \sum_{n=1}^N \gamma(z_{nk})$

Un método muy difundido que aproxima el estimador de máxima verosimilitud de la FDP implícita es el algoritmo de *esperanza y maximización* (EM).

Los modelos mixtos finitos son altamente efectivos cuando una estimación apropiada del número de componentes en la mezcla es muy evidente; sin embargo, cuando se realiza la estimación de una FDP no paramétrica se incrementa el número de componentes arbitrariamente, surgiendo dificultades debido a las características de la función de verosimilitud (Archambeau y Verleysen, 2004).

2.2. Algoritmo EM

El algoritmo de *esperanza y maximización* (EM), fue desarrollado por Dempster, Laird y Rubin (1977), es un método que permite encontrar estimadores máximo verosímiles de los parámetros, cuyo conjunto de datos presenta una distribución implícita, cuando los datos están incompletos o existen datos faltantes.

Existen dos principales aplicaciones del algoritmo EM: El primero se da cuando el conjunto de datos presenta valores faltantes, debido a problemas o limitaciones en el proceso de recojo de información. El segundo ocurre cuando al optimizar la función de verosimilitud se observa que es analíticamente intratable; sin embargo, la función de verosimilitud puede ser simplificada si se asume la existencia de valores para parámetros adicionales pero *ocultos*.

Bajo el supuesto de que el valor faltante contiene información relevante para estimar los parámetros, y estos a su vez permiten encontrar los valores probables de estos datos faltantes, Little y Rubin (1987) sugieren el siguiente esquema para el manejo de datos faltantes, el cual es formalizado por el algoritmo EM:

1. Reemplazar datos faltantes por valores estimados.
2. Estimar parámetros.
3. Estimar los valores faltantes asumiendo que los nuevos parámetros estimados son correctos.
4. Reestimar los parámetros.
5. Repetir los pasos anteriores hasta alcanzar el criterio de convergencia.

Este algoritmo está conformado por dos etapas: La etapa de *esperanza*, E y la etapa de *maximización*, M; estos pasos son frecuentemente fáciles de construir conceptualmente y cada uno tiene una interpretación estadística directa.

Sea $z = (x, y)$ el conjunto completo de datos, donde x representa los datos observados (que han sido generados por alguna distribución) e y , los datos faltantes. Por lo tanto, la función de densidad conjunta dado un conjunto de parámetros θ , está dada por

$$p(Z/\theta) = p(X, Y/\theta) = p(Y/X, \theta)p(X, \theta) \quad (2.6)$$

Así mismo, la función de densidad conjunta frecuentemente surge de la función de densidad marginal $p(X/\theta)$ y el supuesto de variables ocultas y valores estimados (como funciones de densidades mixtas); en otros casos, de la relación conjunta entre los valores observados y faltantes (como los valores faltantes en muestras de una distribución específica) (Bilmes, 1998).

Como la información faltante, y , es desconocida, aleatoria y posiblemente sigue una distribución implícita, entonces se define la función de log-verosimilitud de la siguiente manera:

$$l(\theta/z) = l(\theta/x) + \log p(y/x, \theta)$$

Para la formulación del algoritmo EM, se considera los promedios de la distribución predictiva $\log p(y/x, \theta^i)$, donde θ^i representa un estimado del parámetro desconocido θ , definiéndose de esta manera la función Q , la cual se desea optimizar:

$$Q(\theta, \theta^i) = l(\theta/x) + H(\theta/\theta^i)$$

$$Q(\theta, \theta^i) = E_{y/x, \theta^i}[l(\theta/z)],$$

$$l(\theta/x) = E_{y/x, \theta^i}[l(\theta/x)],$$

$$H(\theta/\theta^i) = E_{y/x, \theta^i}[\log p(y/x, \theta)]$$

donde:

θ^i : Estimaciones de los parámetros actuales

θ : Nuevos parámetros

x : Datos observados

y : Datos no observados

$\log p(y/x, \theta)$: Log verosimilitud de los datos faltantes dado los valores observados.

Implementación del Algoritmo EM

Se inicia seleccionando un $\theta^{(0)}$ arbitrario

Paso E:

Calcula la esperanza condicional del logaritmo de función de verosimilitud de los datos completos respecto a los datos observados y los estimadores de los parámetros actuales:

$$Q(\theta, \theta^i) = E_{y/x, \theta^i}[l(\theta/z)]$$

Paso M:

El paso M del algoritmo maximiza la esperanza calculada en el primer paso

$$\Theta = \max Q(\theta, \theta^i)$$

Este proceso se realiza hasta alcanzar su convergencia hacia el estimador de máxima verosimilitud porque bajo ciertas condiciones generales cada iteración incrementa la verosimilitud de los parámetros dado los valores observados. (Little y Rubin, 1987).

$$Q(\theta^{(i+1)}, \theta^i) \geq Q(\theta, \theta^i)$$

Para todo $\theta \in \Theta$.

De esta manera, se espera que el proceso de estimación converge a la mejor solución global; sin embargo, a veces el algoritmo puede converger a un máximo local, la misma que representa la mejor solución en una vecindad del espacio paramétrico. Para garantizar la obtención de *máximos globales*, se debe realizar el algoritmo de estimación en reiteradas ocasiones con distintos valores iniciales y verificar que se alcance igual solución; si las soluciones son diferentes, elegir aquella que proporcione la mejor bondad de ajuste. Si se encuentra el mismo valor para el estadístico de bondad de ajuste, pero los estimadores son diferentes, existe un problema de *identificación* del modelo; no existe solución única, lo cual ocurre generalmente cuando se considera una cota demasiado grande, esto induce a obtener diferentes soluciones en sucesivas aplicaciones del algoritmo, dependiendo de los valores iniciales asignados a los parámetros, aun cuando la solución sea única (Sepúlveda, Vicente y Galindo, 2008).

Ejemplo

Considerando el ejemplo realizado por Singh (2005), quién aplicó el algoritmo al modelo mixto binomial.

Sean dos monedas con probabilidades desconocidas de obtener *cara*, denotado por p y q respectivamente. Se selecciona la primera moneda con probabilidad π y la segunda, con probabilidad $1 - \pi$. Se lanzó la moneda seleccionada doce veces obteniendo como resultado $x = \{1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1\}$, donde *cara*=1 y *sello*=0. Sea $Y_i \in \{0, 1\}$, la variable latente que se añadió al análisis, que representa lo que se obtuvo en cada lanzamiento, entonces los parámetros que se desean estimar son $\Theta = (p, q, \pi)$.

Paso E:

$$\begin{aligned} Q(\Theta, \Theta^i) &= E \left[\log \prod_{j=1}^n [\pi p^{x_j} (1-p)^{(1-x_j)}]^{y_j} [(1-\pi) q^{x_j} (1-q)^{(1-x_j)}]^{(1-y_j)} \right] \\ &= \sum_{j=1}^n E[y_j/x_j, \Theta^i] [\log \pi + x_j \log p + (1-x_j) \log(1-p)] \\ &\quad + (1 - E[y_j/x_j, \Theta^i]) [\log(1-\pi) + x_j \log q + (1-x_j) \log(1-q)] \end{aligned}$$

Calculando $E[y_j/x_j, \Theta^i]$ se obtiene

$$\begin{aligned} \mu_j^i &= E[y_j/x_j, \Theta^i] = p(y_j = 1/x_j, \Theta^i) \\ &= \frac{p(x_j/y_j, \Theta^i) p(y_j = 1 / \Theta^i)}{p(x_j/\Theta^i)} \end{aligned}$$

$$= \frac{\pi [p^{(i-1)}]^{x_j} [1 - p^{(i-1)}]^{(1-x_j)}}{\pi^i [p^i]^{x_j} [1 - p^i]^{(1-x_j)} + (1 - \pi^i) [q^i]^{x_j} [1 - q^i]^{(1-x_j)}}$$

Paso M:

Maximizando $Q(\theta, \theta^i)$ se obtiene las siguientes ecuaciones:

$$\frac{\partial Q(\theta, \theta^i)}{\partial \pi} = 0 \Rightarrow \pi^{i+1} = \frac{1}{n} \sum_j \mu_j^i$$

$$\frac{\partial Q(\theta, \theta^i)}{\partial p} = 0 \Rightarrow p^{i+1} = \frac{\sum_j \mu_j^i x_j}{\sum_j \mu_j^i}$$

$$\frac{\partial Q(\theta, \theta^i)}{\partial q} = 0 \Rightarrow q^{i+1} = \frac{\sum_j (1 - \mu_j^i) x_j}{\sum_j (1 - \mu_j^i)}$$

2.3. Distribución Logística

La distribución Logística es una función de densidad de probabilidad continua, la cual es simétrica y unimodal. La distribución logística ha sido utilizada para modelos de crecimiento y es usado también en modelos de regresión.

Función de densidad de probabilidad logística

La función de densidad de probabilidad logística es simétrica y se caracteriza por presentar dos parámetros: de localización (μ) y escala (σ)

$$f(x) = \frac{1}{\sigma} * \frac{e^{-(x-\mu)/\sigma}}{(1 + e^{-(x-\mu)/\sigma})^2}$$

$$-\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

La distribución acumulada de la logística es la famosa curva S en la cual se basa la regresión logística y está dada por:

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}}$$

Respecto a los momentos, se observa que el valor esperado es igual al parámetro de localización:

$$E(x) = \mu$$

y la varianza de la distribución es igual a:

$$Var(x) = \frac{1}{3}(\pi\sigma)^2$$

Considerando una distribución logística con parámetros $\mu = 0$ y $\sigma = 1$, se observa que la varianza es $\pi^2/3$ (Snijders y Bosker, 1999).

CAPÍTULO 3

Modelo de Clases Latentes

3.1. Introducción

En diversas áreas del conocimiento, especialmente en el campo de las ciencias sociales, muchos conceptos no pueden ser directamente observados o cuantificados como las variables que frecuentemente son utilizadas como medidas indirectas de conceptos que son medidos (Vermunt & Magidson, 2002).

Según Bartholomew y Knoot (1999) la selección del modelo dependerá de la escala de medición tanto de las variables observadas como el de las variables latentes. En el Cuadro N°1.1 se presenta la clasificación de estos modelos.

Cuadro 3.1. Clasificación de modelos con variables latentes

		Variables manifiestas	
		Continuas	Categóricas
Variables latentes	Continuas	Análisis factorial	Análisis de rasgo latente
	Categóricas	Análisis de perfil latente	Análisis de clases latentes

En el análisis factorial, las variables observadas continuas son utilizadas como indicadores de una o más variables latentes continuas. En los modelos de rasgos latentes, la variable latente es continua; mientras que, las variables observadas son categóricas. Cuando ambas variables, tanto las variables latentes como las variables observadas son categóricas; se considera a los modelos de clases latentes.

El Análisis de Clases Latentes (ACL) fue propuesto por Lazarsfeld (1950), quién uso la técnica como una herramienta para construir una tipología en el análisis de un conjunto de variables dicotómicas, años más tarde Lazarsfeld y Henry (1968) continuaron utilizando el modelo, para segmentar a los datos observados.

Goodman (1974) hizo el modelo aplicable en la práctica, desarrollando un algoritmo para obtener estimadores de MV de los parámetros; propuso también una extensión del modelo para variables politómicas y múltiples variables latentes e hizo un trabajo importante en el problema de identificación del modelo.

El MCL es miembro de los modelos de mezclas finitas (Vermunt, 2008). El modelo mixto es analizado bajo el supuesto de que los datos provienen de una fuente con varias subpoblaciones, cada una de las cuales es modelada de forma separada; siendo el total de datos representado por una mezcla de estas subpoblaciones y el modelo resultante, una mezcla de componentes finitos. La forma general de un modelo mixto con K componentes es

$$f(X_i) = \sum_{k=1}^K p_k f_k(X_i/\theta_k)$$

donde:

X_i : i -ésima observación, $i=1, \dots$

p_k : Proporción mixta en el k -ésimo grupo.

$f_k(.|\theta_k)$: Función de densidad del k -ésimo grupo con parámetro θ_k .

Varias extensiones importantes de este modelo son similares a aquellos que presentan más de una variable latente o modelos que cuentan con covariables (Goodman, 1974); modelos para análisis de grupos múltiples (Clogg y Goodman, 1984) y modelos con efectos directos entre las variables observadas (Hagenaars, 1988).

La idea básica del ACL se presenta cuando las variables observadas sirven para definir o medir la variable latente (Vermunt y Magidson, 2002).

Sea $X' = (X_1, X_2, \dots, X_K)$ el vector de K variables observadas e $Y' = (Y_1, Y_2, \dots, Y_q)$ el vector de q variables latentes del modelo para una muestra de n individuos con $q < K$, la distribución condicional conjunta denotado por $g(x/y)$ de las variables observadas X_1, X_2, \dots, X_K ; sobre las variables latentes,

$$g(x/y) = P(X = x/Y = y) \quad (3.1)$$

La función de densidad de la variable latente Y se denota como $h(y)$, luego en (3.1)

$$P(X = x) = f(x) = \int g(x/y)h(y)dy \quad (3.2)$$

Independencia local, dado un valor fijo de la variable latente, las variables observadas son mutuamente independientes, por lo que la distribución condicional se descompone en el producto de probabilidades condicionales marginales,

$$g(x/y) = \prod_{i=1}^p g_i(x_i/y) \quad (3.3)$$

Reemplazando (3.3) en (3.2) se tiene:

$$f(x) = \int \prod_{i=1}^p g_i(x_i/y) h(y) dy \quad (3.4)$$

3.2. Planteamiento del Modelo de Clases Latentes

El MCL se analiza bajo el supuesto de independencia local (Lazarsfeld y Henry, 1968), el cual determina si las relaciones de un conjunto de variables observadas son espurias; es decir, las relaciones existentes desaparecen cuando se incorpora una variable no observada o latente (Mc Cutcheon, 1987).

Notación

Sea la matriz de información que contiene K variables categóricas (nominales u ordinales) sobre una muestra de n individuos,

$$\mathbf{X} = \begin{matrix} & X_1 & X_2 & \dots & X_K \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n \end{matrix} & \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nK} \end{bmatrix} \end{matrix}$$

Figura 3.1. **Matriz de datos**

Cada fila de la matriz representa un patrón de respuesta $x' = (x_1, \dots, x_K)$; con respecto a las variables, estas pueden ser binarias o politómicas con S_k número de niveles, donde $x_{ik} = 1, 2, \dots, S_k; \forall k = 1, 2, \dots, K$.

Mediante el MCL se estudian las relaciones que existe entre las variables observadas considerando variables latentes en el análisis; por tal motivo, la tabla que sólo contiene variables observadas se le denomina *incompleta*; por el contrario, la tabla que contiene ambos tipos de variables (observables y latentes) se considera como *completa*.

Para la presentación del MCL en esta sección, se considera una variable latente Y con T categorías o clases latentes. Las K variables observadas se consideran indicadoras de la variable latente Y .

Sea $P(X = s)$, la función de densidad conjunta de las variables observadas $X' = (X_1, \dots, X_K)$, donde el vector $s = (s_1, \dots, s_K)$ denota un determinado patrón de respuesta en el cual cada s_k toma diferentes valores dependiendo de las categorías de la correspondiente variable observada; este vector de respuestas se presenta en función de la probabilidad de que el individuo pertenezca a una clase de la variable latente y la probabilidad de que obtenga un valor en cada variable observada dado la pertenencia a la clase latente t , $t=1, 2, \dots, T$.

Las variables observadas conforman una tabla de contingencia múltiple con $\prod_{k=1}^K S_k$ patrones de respuesta, donde cada X_k contiene S_k categorías.

Parametrizaciones del MCL

El MCL puede ser parametrizado de dos formas (parametrización probabilística y parametrización log-lineal); demostrándose así, la equivalencias de ambas y como se manifiestan las propiedades comunes (independencia local) en cada una de ellas.

El incumplimiento del principio de *Independencia local* trae como consecuencia una falta de ajuste del modelo (Vermunt & Magidson, 2002).

1. Parametrización Probabilística

La parametrización probabilística del MCL básico se caracteriza por considerar en el análisis dos tipos de variables categóricas (observadas y latentes) y dos tipos de parámetros (clases latentes y probabilidades condicionales). Por el principio de *independencia local*, la densidad condicional, de un determinado individuo con un patrón de respuesta s pertenezca a la clase latente t , está dada por,

$$P(\mathbf{X} = \mathbf{s}/Y = t) = \prod_{k=1}^K P(X_k = s_k/Y = t) \quad (3.5)$$

donde:

$P(\mathbf{X} = \mathbf{s}/Y = t)$: Probabilidad condicional de individuos que se encuentran en cada clase latente t , $t = 1, 2, \dots, T$.

$P(X_k = s_k/Y = t)$: Probabilidad condicional de cada una de las variables observadas dentro de cada clase latente, $k = 1, 2, \dots, K$ y $t = 1, 2, \dots, T$.

La distribución conjunta de \mathbf{X} e Y , $P(\mathbf{X} = \mathbf{s}, Y = t)$ es de la siguiente forma,

$$P(\mathbf{X} = \mathbf{s}, Y = t) = P(Y = t)P(\mathbf{X} = \mathbf{s}/Y = t) \quad (3.6)$$

$P(Y = t)$ representa la probabilidad *a priori*, la cual indica la proporción de elementos que se encuentran en la clase latente t .

De (3.5) y (3.6), el Modelo de Clases Latentes puede ser escrito de la siguiente forma

$$P(\mathbf{X} = \mathbf{s}) = \sum_{t=1}^T P(Y = t) \prod_{k=1}^K P(X_k = s_k/Y = t) \quad (3.7)$$

Esta expresión implica que la población puede ser dividida en T clases latentes exhaustivas y exclusivas, entonces la probabilidad conjunta de las variables observadas se obtienen sumando las dimensiones latentes. Por lo tanto, esta expresión implica la existencia de una variable latente.

Reemplazando (3.6) en (3.7), el MCL se puede representar también,

$$P(\mathbf{X} = \mathbf{s}) = \sum_{t=1}^T P(Y = t)P(\mathbf{X} = \mathbf{s}/Y = t)$$

$$P(\mathbf{X} = \mathbf{s}) = \sum_{t=1}^T P(Y = t) \prod_{k=1}^K P(X_k = s_k/Y = t) \quad (3.8)$$

Así mismo los parámetros están sujetos a las siguientes restricciones

$$\sum_{t=1}^T P(Y = t) = 1, \quad \sum_{s_k=1}^{S_k} P(X_k = s_k/Y = t) = 1; s_k = 1, \dots, S_k \text{ y } t = 1, \dots, T$$

donde:

$P(Y = t)$: Probabilidad a priori.

$P(X_k = s_k/Y = t)$: Probabilidad condicional de la variable observada k dada la clase latente t , $k = 1, \dots, K$ y $t = 1, \dots, T$.

La representación gráfica de un MCL con una variable latente Y y K variables observadas se presentan en la Figura 3.2, observándose que no están directamente conectadas; la dirección de las flechas indica que la variable latente Y explica toda posible asociación entre las variables observadas.

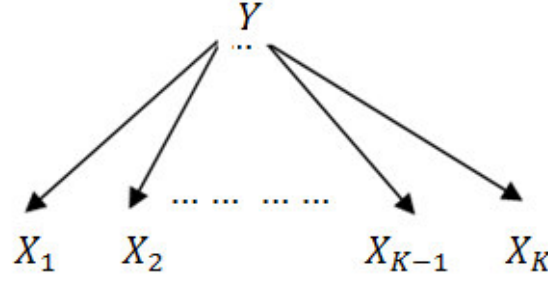


Figura 3.2. **Modelo de Clases Latentes con una variable latente y K variables observadas**

El MCL cuantifica las probabilidades de pertenencia de cada individuo en las T clases latentes, y asignará a cada uno de estos a la clase latente donde ha obtenido una mayor puntuación.

Una vez realizado el análisis se puede considerar la distribución condicional de Y dado X , $P(Y = t/X = s)$, conocida también como distribución *a posteriori*,

$$P(Y = t/X = s) = \frac{P(X = s, Y = t)}{P(X = s)} = \frac{P(Y = t)P(X = s|Y = t)}{P(X = s)} \quad (3.9)$$

2. Parametrización Log-lineal

El MCL se puede representar también como un modelo log-lineal (Goodman, 1974; Haberman, 1979); al igual que en la parametrización probabilística del MCL, la parametrización log-lineal también manifiesta la independencia local entre las variables observadas (McCutcheon, 1987).

El MCL log-lineal para tres variables observadas (X_1, X_2, X_3) y una variable latente Y se expresa como,

$$\ln(f_{x_1 x_2 x_3 t}^{X_1 X_2 X_3 Y}) = \beta + \beta_t^Y + \beta_{x_1}^{X_1} + \beta_{x_2}^{X_2} + \beta_{x_3}^{X_3} + \beta_{x_1 t}^{X_1 Y} + \beta_{x_2 t}^{X_2 Y} + \beta_{x_3 t}^{X_3 Y} \quad (3.10)$$

donde:

$f_{x_1 x_2 x_3 t}^{X_1 X_2 X_3 Y}$: Frecuencia esperada para la celda que pertenece a la categoría

x_k de X_k y la clase latente t de Y . $k = 1, 2, 3$.

$x_k = 1, 2, \dots, S_k$ y $t = 1, 2, \dots, T$

$\beta_i^{X_1}$: Efecto principal de la variable observada X_1 .

$\beta_j^{X_2}$: Efecto principal de la variable observada X_2 .

$\beta_k^{X_3}$: Efecto principal de la variable observada X_3 .

β_t^Y : Efecto principal de la variable latente Y .

$\beta_{jk}^{X_1 Y}$: Efecto de interacción entre X_1 y Y .

$\beta_{ik}^{X_2 Y}$: Efecto de interacción entre X_2 y Y .

$\beta_{jk}^{X_3 Y}$: Efecto de interacción entre X_3 y Y .

Así como en los modelos log-lineales, en este tipo de parametrización, los parámetros también están sujetos a las siguientes restricciones

$$\sum_{t=1}^T \beta_t^Y = \sum_{x_k=1}^{S_k} \beta_{x_k}^{X_k} = \sum_{x_k=1}^{S_k} \beta_{x_k t}^{X_k Y} = \sum_{t=1}^T \beta_{x_k t}^{X_k Y} = 0$$

$$x_k = 1, \dots, S_k \text{ y } t = 1, \dots, T$$

Al considerar al MCL como un modelo log-lineal, se puede obtener una parametrización *logit* para las probabilidades condicionales,

$$P(X_k = s_k / Y = t) = \frac{e^{(\beta_{s_k} + \beta_{s_k t})}}{\sum_{r=1}^{S_k} e^{(\beta_r + \beta_{rt})}} \quad (3.11)$$

$$k = 1, 2, \dots, K; x_k = 1, 2, \dots, S_k \text{ y } t = 1, 2, \dots, T.$$

Es importante observar que el MCL desarrollado por cualquiera de las dos parametrizaciones, estima la misma cantidad de parámetros con valores esperados idénticos; esta equivalencia se da principalmente por las probabilidades condicionales de ambas parametrizaciones (Haberman, 1979).

3.3. Estimación del Modelo de Clases Latentes

Para estimar los parámetros del MCL, existen dos alternativas principales, los algoritmos *Esperanza y Maximización* (EM) y *Newton Raphson* (NR), ambos con una aproximación iterativa de la estimación de máxima verosimilitud. Como procedimientos iterativos, estos algoritmos empiezan con un conjunto de *valores iniciales* y proceden con un conjunto de pasos iterativos de estimación y reestimación de los parámetros mediante un criterio de convergencia.

Algoritmo EM

El algoritmo EM viene a ser una de las aproximaciones más utilizadas en la estimación de los MCL. En este algoritmo los *datos faltantes* representan las clases a las que pertenecen los individuos y los *estimadores* de estos valores son las probabilidades *a posteriori*, $P(Y = t/X = s)$.

Paso E: Cálculo de las probabilidades *a posteriori*, las cuales representan los valores esperados dado los datos observados y los parámetros *actuales* del modelo.

Paso M: En el MCL, para el caso de variables observadas binarias, el logaritmo de la función de verosimilitud está dado por

$$\begin{aligned}
Ln(L) &= \sum_{s=1}^R f(\mathbf{X} = \mathbf{s}) \\
&\quad Ln \left\{ \sum_{t=1}^T P(Y = t) \prod_{k=1}^K [P(X_k = s_k | Y = t)]^{s_k} [1 - P(X_k = s_k | Y = t)]^{1-s_k} \right\} \\
&= \sum_{s=1}^R f(\mathbf{X} = \mathbf{s}) Ln[P(\mathbf{X} = \mathbf{s})]
\end{aligned} \tag{3.12}$$

donde:

$f(\mathbf{X} = \mathbf{s})$: Frecuencia observada del patrón de respuesta \mathbf{s} .

$P(\mathbf{X} = \mathbf{s})$: Probabilidad del patrón de respuesta \mathbf{s} .

R : Número posible de patrones de respuesta (2^K).

Este proceso iterativo de estimación finalizará cuando se alcance una diferencia mínima denominada nivel de tolerancia, entre la estimación máximo verosímil actual y la estimación máximo verosímil inmediatamente anterior.

Las dos principales desventajas mencionadas frecuentemente del algoritmo EM son

- El gran número de iteraciones que puede necesitar esta aproximación para alcanzar la solución final.
- El algoritmo no provee estimadores directamente de los errores estándar.

Actualmente los programas computacionales han reducido el primero de estos problemas (McCutcheon, 1987).

Algoritmo NR

El método NR también utiliza una aproximación iterativa para producir estimadores de máxima verosimilitud de los parámetros de los MCL. Este

método tiene la ventaja de ser relativamente rápido y producir errores estándar de las estimaciones de parámetros. Las desventajas que presenta este método son la necesidad de invertir la matriz Hessiana en cada iteración y el requerimiento de que las estimaciones iniciales deben estar cerca de las estimaciones finales, permitiendo de esta manera la convergencia a la solución final o que la matriz pueda ser definida negativa y así no pueda ser invertida.

El método NR comienza con un conjunto de valores iniciales de los parámetros (θ) y mejora estos valores modificándolos por el producto de la inversa de la matriz Hessiana (H) y el vector gradiente (g) de derivadas de la función log-verosimilitud.

$$\theta_{i+1} = \theta_i - H_i^{-1} g_i$$

Los parámetros estimados son actualizados al final de cada iteración y la convergencia de este método va a ser bastante rápido en la medida que los parámetros estimados (θ) inicialmente sean cercanos a los estimadores de MV (McCutcheon, 1987).

Los métodos de estimación EM y NR convergen a un óptimo local y pueden no representar al verdadero máximo global de la función de verosimilitud; por este motivo, se sugiere repetir el procedimiento usando diferentes valores iniciales.

Una vez solucionado el problema de estimaciones de los parámetros del modelo, se debe evaluar el modelo. Para tal fin se describirá los estadísticos de bondad de ajuste en la siguiente sección.

3.4. Bondad de ajuste del Modelo de Clases Latentes

Los estadísticos de bondad de ajuste se utilizan para decidir si una muestra observada con patrones de respuesta específicos, puede ser considerada como

un conjunto de realizaciones independientes de una función de distribución dada.

La evaluación de los MCL se realiza por criterios más o menos estándar, comparando las frecuencias observadas para cada patrón de respuesta f_x , con las frecuencias esperadas dado por los parámetros estimados del MCL el modelo, \hat{f}_x . En esta sección se analizará cuatro de estos criterios, el estadístico ji-cuadrado de Pearson (χ^2), la Razón de verosimilitud (L^2), el estadístico de Cressie-Read (CR) y el estadístico de Freeman-Tukey (FT^2). Para el cálculo de estos estadísticos es importante definir las frecuencias esperadas.

Las probabilidades para un patrón de respuesta $x = (x_1, \dots, x_K)$,

$$p(X = s) = \sum_{t=1}^T p(Y = t) \prod_{k=1}^K p(X_k = s_k / Y = t)$$

Entonces, las frecuencias esperadas son,

$$\hat{f}_x(x) = n \sum_{t=1}^T p(Y = t) \prod_{k=1}^K p(X_k = s_k / Y = t)$$

Cuadro 3.2. Estadísticos de bondad de ajuste

Estadístico	Cálculo
χ^2	$\sum_x \frac{(f_x - \hat{f}_x)^2}{\hat{f}_x}$
L^2	$2 \sum_x f_x \ln \left(\frac{f_x}{\hat{f}_x} \right)$
CR	$\frac{2}{\lambda(\lambda+1)} \sum_x f_x \left[\left(\frac{f_x}{\hat{f}_x} \right)^\lambda - 1 \right], \quad \lambda \neq 0$
FT^2	$4 \sum_x \left(\sqrt{f_x} - \sqrt{\hat{f}_x} \right)^2$

Notación:

f_x : Frecuencia observada.

\hat{f}_x : Frecuencia esperada.

Los estadísticos χ^2 de Pearson y razón de verosimilitud L^2 siguen una distribución ji-cuadrada con grados de libertad igual a,

$$\prod_{k=1}^K S_k - m - 1 \quad (3.13)$$

m representa al número de parámetros independientes a estimar en el modelo; así mismo, se estiman $T - 1$ probabilidades para las clases latentes y $[\sum_{k=1}^K (S_k - 1)]$ probabilidades condicionales para cada clase latente; entonces,

$$m = (T - 1) + T * \sum_{k=1}^K (S_k - 1) = T * \left(\sum_{k=1}^K (S_k - 1) + 1 \right) - 1 \quad (3.14)$$

Reemplazando (3.13) en (3.14), los grados de libertad (gl) es igual a

$$gl = \prod_{k=1}^K S_k - T(\sum_{k=1}^K (S_k - 1) + 1)$$

Collins (1993) demostró mediante un estudio de simulación que en el Análisis de Clases Latentes, el estadístico χ^2 es más apropiado que G^2 ; así mismo, ambos estadísticos poseen ciertas limitaciones:

- Tienden a ser conservadoras cuando el tamaño de muestra es grande; es decir, es difícil rechazar la significancia, incluso para modelos que estimen un mínimo número de parámetros cuando el tamaño de la muestra es grande.
- Los MCL pueden requerir la estimación de un número bastante grande de parámetros, incluso para los modelos con pequeños tamaños de muestra.

3.5. Criterios de selección del Modelo de Clases Latentes

Cuando se tiene una serie de p modelos M_1, M_2, \dots, M_p con parámetros k_1, k_2, \dots, k_p , el objetivo es descubrir el modelo que proporcione mayor información de los datos; por este motivo, se espera que las frecuencias

estimadas sean parecidas a las observadas y a su vez, que este modelo sea parsimonioso (menor número de parámetros).

Una metodología para compararlos corresponde a la función de máxima verosimilitud. La MV permite seleccionar el modelo que realiza el mejor ajuste de los datos pero no penaliza su complejidad, lo que si sucede cuando se emplean medidas de contraste como el Criterio de información de Akaike (AIC) y el Criterio de información bayesiano (BIC). Ambos criterios hacen uso del logaritmo de MV,

$$AIC = -2\ln(\iota) + 2m$$

$$BIC = -2\ln(\iota) + m * \ln(n)$$

donde:

ι : Valor de la función de verosimilitud.

n : Tamaño de muestra.

m : Número de parámetros independientes estimados.

Los modelos con mínimos valores del AIC y BIC son mejores que aquellos modelos que presentan valores más altos para estos criterios.

3.6. Modelo de Clases Latentes Multi-grupos

El MCL multi-grupos es una extensión del MCL estándar que ha sido utilizado para comparar modelos a través de los grupos; es decir, estimar un MCL con t clases para cada grupo (Clogg y Goddman, 1984).

El MCL multi-grupos considera tres tipos de variables categóricas: variables observadas, una variable latente que es explicada a través de las variables observadas y una variable *grupo*.

Definición del modelo

Sea X_g una variable grupo e Y , la variable latente. El modelo de clases latentes multi-grupos para tres variables observadas (X_1 , X_2 y X_3) se define de la siguiente manera:

$$\pi_{ijk/g}^{X_1X_2X_3/X_g} = \sum_{t=1}^T \pi_{t/g}^{Y/X_g} \pi_{i/tg}^{X_1/YX_g} \pi_{j/tg}^{X_2/YX_g} \pi_{k/tg}^{X_3/YX_g}$$

$$\sum_{t=1}^T \pi_{t/g}^{Y/X_g} = 1; \sum_{i=1}^{n_1} \pi_{i/tg}^{X_1/YX_g} = 1; \sum_{j=1}^{n_2} \pi_{j/tg}^{X_2/YX_g} = 1; \sum_{k=1}^{n_3} \pi_{k/tg}^{X_3/YX_g} = 1$$

$$i = 1, 2, \dots, n_1; j = 1, 2, \dots, n_2; k = 1, 2, \dots, n_3; t = 1, 2, \dots, T; g = 1, 2, \dots, G.$$

donde:

$$\pi_{ijk/g}^{X_1X_2X_3/X_g} = P(X_1 = i, X_2 = j, X_3 = k / X_g = g)$$

$$i = 1, 2, \dots, n_1; j = 1, 2, \dots, n_2; k = 1, 2, \dots, n_3; g = 1, 2, \dots, G.$$

$$\pi_{t/g}^{Y/X_g} = P(Y = t / X_g = g) \quad t = 1, 2, \dots, T; g = 1, 2, \dots, G.$$

$$\pi_{i/tg}^{X_1/YX_g} = P(X_1 = i / X_g = g) \quad i = 1, 2, \dots, n_1; g = 1, 2, \dots, G.$$

$$\pi_{j/tg}^{X_2/YX_g} = P(X_2 = j / X_g = g) \quad j = 1, 2, \dots, n_2; g = 1, 2, \dots, G.$$

$$\pi_{k/tg}^{X_3/YX_g} = P(X_3 = k / X_g = g) \quad k = 1, 2, \dots, n_3; g = 1, 2, \dots, G.$$

En este modelo, las probabilidades varían a través de los grupos. Para la variable observada X_1 , se cumple la siguiente propiedad:

$$\pi_{i/tg}^{X_1/YX_g} = \pi_{i/t}^{X_1/Y}$$

Parametrización

Las probabilidades de respuesta condicional de la parametrización probabilística se obtienen en términos log-lineales (Haberman, 1979). La

parametrización de la variable X_1 respecto a la variable latente y variable grupo

$\pi_{x_1/tg}^{X_1/YX_g}$, es el siguiente:

$$\pi_{x_1/tg}^{X_1/YX_g} = \frac{\exp(\gamma_i^{X_1} + \gamma_{it}^{X_1Y} + \gamma_{ig}^{X_1X_g} + \gamma_{itg}^{X_1YX_g})}{\sum_{x_1=1}^I \exp(\gamma_i^{X_1} + \gamma_{it}^{X_1Y} + \gamma_{ig}^{X_1X_g} + \gamma_{itg}^{X_1YX_g})}$$

donde:

$\gamma_i^{X_1}$: Efecto de la variable observada X_1 .

$\gamma_{it}^{X_1Y}$: Efecto de interacción entre X_1 e Y .

$\gamma_{ig}^{X_1X_g}$: Efecto directo de la variable grupo X_g en la variable observada X_1 .

$\gamma_{itg}^{X_1YX_g}$: Efecto de interacción entre X_1 , Y y X_g .

El efecto de interacción entre X_1 , Y y X_g , describe la relación que existe entre la variable observada y la variable latente, la cual se modifica por la pertenencia al grupo y el efecto directo $\gamma_{ig}^{X_1X_g}$ se presenta cuando la variable grupo X_g , influye en las variables observadas independientemente de la variable latente Y .

De forma similar se puede definir las probabilidades de pertenencia a la clase

$\pi_{t/g}^{Y/X_g}$, en términos de parámetros log-lineales:

$$\pi_{t/g}^{Y/X_g} = \frac{\exp(\gamma_t^Y + \gamma_{tg}^{YX_g})}{\sum_{t=1}^T \exp(\gamma_t^Y + \gamma_{tg}^{YX_g})}$$

donde:

γ_t^Y : Efecto de la variable latente Y .

$\gamma_{tg}^{YX_g}$: Efecto de interacción entre Y y X_g .

CAPÍTULO 4

Modelo de Clases Latentes Multinivel (MCLM)

4.1. Introducción

El Análisis de Clases Latentes (ACL), es un método ampliamente utilizado con el objetivo de construir tipologías a partir de variables categóricas observadas. El ACL asume que las observaciones son independientes, el cual es infringido frecuentemente; dado que en la práctica se observa una interacción entre los individuos y el contexto al cual pertenecen; como estudiantes dentro de sus colegios, empleados dentro de sus compañías, etc.; estas estructuras de datos anidados requieren de Modelos Multinivel. Vermunt (2003) y Asparouhov y Muthen (2008) presentan una estructura para evaluar MCL con datos anidados; denominado Modelo de Clases Latentes Multinivel (MCLM).

En el MCL estándar, se asume que los parámetros del modelo son iguales para todos los individuos (unidades del primer nivel); a diferencia del MCLM, donde ciertos parámetros difieren a través de los grupos (unidades del segundo nivel); tales diferencias pueden ser modeladas incluyendo variables *dummies* en el modelo, como un análisis de clases latentes multi-grupos (Clogg y Goodman, 1984) basado en un enfoque de efectos fijos, donde cada grupo tiene su propio conjunto de probabilidades de pertenencia a cada clase, sin embargo, los parámetros del modelo no pueden ser estimados bajo una estructura multinivel (Vermunt, 2008). Una alternativa a este modelo es un enfoque de efectos aleatorios, bajo el supuesto de que los coeficientes de los grupos específicos siguen una distribución particular, el modelo es analizado mediante un enfoque no paramétrico.

El MCLM es similar al Modelo de Regresión Logística de coeficientes aleatorios; sin embargo, existe una diferencia principal entre estos dos modelos, en el MCLM, la variable dependiente no es directamente observada, es una variable latente que es explicada por las variables observadas.

4.2. Definición del modelo

Sea X_{ijk} la respuesta de la observación i del primer nivel dentro del grupo j de la variable observada k . El número total de unidades dentro de cada grupo es denotado por n_j ; $j = 1, 2, \dots, J$ y el número de variables observadas por K . Una respuesta particular de la variable k es denotada por s_k y el número total de categorías, por S_k . La variable de clase latente se encuentra definida por Y_{ij} ; mientras que, una clase latente en particular, por t , $t = 1, 2, \dots, T$. Sea \mathbf{X}_{ij} el vector de respuestas del individuo i dentro del grupo j y \mathbf{s} , un posible patrón de respuesta.

La probabilidad definida mediante el MCL estándar mostrada en la ecuación (3.8), presenta la siguiente forma:

$$P(\mathbf{X}_{ij} = \mathbf{s}) = \sum_{t=1}^T P(Y_{ij} = t) P(\mathbf{X}_{ij} = \mathbf{s} / Y_{ij} = t) \quad (4.1)$$

donde:

$P(\mathbf{X}_{ij} = \mathbf{s})$: Probabilidad de observar un patrón de respuestas específico.

$P(Y_{ij} = t)$: Probabilidad de que individuo i dentro del grupo j pertenezca a la clase latente t .

$P(\mathbf{X}_{ij} = \mathbf{s} / Y_{ij} = t)$: Probabilidad condicional de observar un patrón de respuestas \mathbf{s} dada la clase específica t .

Bajo el supuesto de independencia local del MCL, dada la pertenencia a la clase en la que se encuentra, se asume que los X_{ijk} son independientes. Entonces la ecuación (4.1) se puede expresar como:

$$P(\mathbf{X}_{ij} = \mathbf{s}) = \sum_{t=1}^T P(Y_{ij} = t) \prod_{k=1}^K P(X_{ijk} = s_k | Y_{ij} = t) \quad (4.2)$$

$P(X_{ijk} = s_k | Y_{ij} = t)$ denota la probabilidad de la respuesta observada s_k en el ítem k , dado que la persona o individuo pertenece a la clase latente t ; así mismo, cada una de estas probabilidades condicionales son ponderadas con la probabilidad $P(Y_{ij} = t)$ que corresponde a la unidad i en el grupo j , perteneciente a la clase latente t . Marginalizando para el grupo j , se obtiene:

$$P(\mathbf{X}_j = \mathbf{s}) = \prod_{i=1}^{n_j} \sum_{t=1}^T P(Y_{ij} = t) \prod_{k=1}^K P(X_{ijk} = s_k | Y_{ij} = t) \quad (4.3)$$

Para calcular $P(Y_{ij} = t)$ y $P(X_{ijk} = s_k | Y_{ij} = t)$ se consideró una de las propiedades del modelo de clases latentes:

$$P(\mathbf{X}_{ij} = \mathbf{s}, Y_{ij} = t) = P(Y_{ij} = t) * P(X_{ij1} = s_1 | Y_{ij} = t) \dots$$

$$P(X_{ij1} = s_k | Y_{ij} = t)$$

$$\begin{aligned} \log[P(\mathbf{X}_{ij} = \mathbf{s}, Y_{ij} = t)] &= \log[P(Y_{ij} = t)] + \log[P(X_{ij1} = s_1 | Y_{ij} = t)] + \\ &\dots + \log[P(X_{ij1} = s_k | Y_{ij} = t)] \end{aligned}$$

$$\log[P(\mathbf{X}_{ij} = \mathbf{s}, Y_{ij} = t)] = \gamma_t + \beta_{s_1 t}^{X_{ij1} Y_{ij}} + \beta_{s_2 t}^{X_{ij2} Y_{ij}} + \dots + \beta_{s_k t}^{X_{ijk} Y_{ij}} \quad (4.4)$$

Sea L_t , la transformación logit generalizada que toma como referencia la última categoría T , la cual se define como:

$$L_t = \log \left[\frac{P(Y_{ij} = t)}{P(Y_{ij} = T)} \right]$$

$\forall t = 1, \dots, T.$

Por lo tanto, el modelo lineal para cada una de las transformaciones logit generalizadas es de la siguiente forma:

$$L_t = \log \left[\frac{P(Y_{ij} = t)}{P(Y_{ij} = T)} \right] = \gamma_t$$

donde

$$\frac{P(Y_{ij} = t)}{P(Y_{ij} = T)} = \exp(\gamma_t)$$

$$\sum_{t=1}^T \frac{P(Y_{ij} = t)}{P(Y_{ij} = T)} = \sum_{t=1}^T \exp(\gamma_t)$$

$$\frac{\sum_{t=1}^T P(Y_{ij} = t)}{P(Y_{ij} = T)} = \sum_{t=1}^T \exp(\gamma_t)$$

$$\frac{1}{P(Y_{ij} = T)} = \sum_{r=1}^T \exp(\gamma_r)$$

$$P(Y_{ij} = T) = \frac{1}{\sum_{r=1}^T \exp(\gamma_r)}$$

entonces

$$P(Y_{ij} = t) = \frac{\exp(\gamma_t)}{\sum_{r=1}^T \exp(\gamma_r)} \quad (4.5)$$

De forma análoga se calcula el valor de $P(X_{ijk} = s_k / = t)$, utilizando el modelo lineal

$$\log \left[\frac{P(X_{ijk} = s_k / Y_{ij} = t)}{P(X_{ijk} = S_k / = t)} \right] = \beta_{s_k t}^{X_{ijk} Y_{ij}}$$

donde

$$P(X_{ijk} = s_k / Y_{ij} = t) = \frac{\exp(\beta_{s_k t})}{\sum_{r=1}^{S_k} \exp(\beta_{rt})} \quad (4.6)$$

De las ecuaciones (4.5) y (4.6) se observa que los parámetros γ y β no tienen el índice j , lo cual indica que sus valores son asumidos a ser independientes de los grupos al que cada uno pertenece.

Por otro lado, tomando en cuenta una estructura multinivel, Vermunt (2003) obtiene un MCLM más general, el cual asume que todos los parámetros en el modelo representan grupos específicos, por lo que la ecuación (4.3) se expresa como un modelo de regresión logística de dos niveles con un intercepto aleatorio γ_{tj} :

$$\ln P(Y_{ij} = t, \mathbf{X}_{ij} = \mathbf{s}) = \gamma_{tj} + \sum_{s_k=1}^{S_k} \beta_{s_k, t j}^{X_{ijk}, Y_{ij}} + \varepsilon_{ij} \quad (4.7)$$

$$\text{var}(\varepsilon_{ij}) = \sigma^2 \approx \pi^2 / 3 = 3.29$$

donde:

β_{0j} : Intercepto aleatorio.

$\beta_{s_k, t j}^{X_{ijk}, Y_{ij}}$: Efecto de interacción entre X_{ijk} y Y_{ij}

ε_{ij} : Efecto aleatorio

Obteniéndose de esta manera $P(Y_{ij} = t)$ y $P(X_{ijk} = s_k/Y_{ij} = t)$ a partir de la ecuación (4.7)

$$P(Y_{ij} = t) = \frac{\exp(\gamma_{tj})}{\sum_{r=1}^T \exp(\gamma_{rj})} \quad (4.8)$$

$$P(X_{ijk} = s_k/Y_{ij} = t) = \frac{\exp(\beta_{s_k t j})}{\sum_{r=1}^{S_k} \exp(\beta_{r t j})} \quad (4.9)$$

Sin otras especificaciones, el modelo definido hasta ahora corresponde a un MCL multi-grupos planteado por Clogg & Goodman (1984), donde cada grupo tiene su propio conjunto de probabilidades de pertenencia y vector de parámetros y es similar a utilizar una aproximación de efectos fijos.

Como todos los parámetros son grupos específicos, se obtiene varios conjuntos de probabilidades de acuerdo al número de unidades del segundo nivel; por lo que, el número de parámetros se incrementa de forma exponencial, una segunda desventaja de este modelo es que no permite distinguir entre los efectos del primer y segundo nivel; lo cual puede representar una fuerte limitación.

Una aproximación de efectos aleatorios no paramétrico en el análisis de clases latentes

Las desventajas mencionadas por el MCL multi-grupos pueden ser evitadas adoptando una aproximación de efectos aleatorios, donde se asume que los coeficientes de los grupos específicos provienen de una distribución en particular y cuyos parámetros deberían ser estimados.

Según Vermunt (2003), la idea básica del MCLM es permitir que ciertos parámetros del modelo difieran aleatoriamente a través de las unidades del nivel 2; por lo que, existe dos formas de poder analizar el modelo, al introducir una variable aleatoria continua o una variable aleatoria discreta en el segundo nivel.

En el primer caso, se asume que los coeficientes aleatorios provienen de una distribución normal, para el caso de T clases, se puede trabajar con una distribución normal $T-1$ dimensional o utilizar una estructura más restringida, el cual se presenta al descomponer el intercepto aleatorio γ_{tj} del modelo presentado en la ecuación (4.7):

$$\gamma_{tj} = \gamma_t + \tau_t u_j \quad (4.10)$$

$$u_j \sim N(0,1)$$

donde:

γ_{tj} : Intercepto aleatorio.

γ_t : Media de la clase latente t .

τ_t : Desviación estándar de la clase latente t .

u_j : Efecto aleatorio.

Bajo el supuesto implícito que existe de una alta correlación entre las componentes aleatorias de cada γ_{tj} en el modelo de regresión de dos niveles con intercepto aleatorio.

Nota: En modelos de efectos aleatorios, la *correlación intraclase* es de mucha utilidad; esta se define como una proporción entre la varianza explicada por las unidades del segundo nivel y la varianza total, la cual equivale a la suma de las varianzas del primer y segundo nivel. Hedeker (1999) propuso como calcular la *correlación intraclase* en modelos de regresión logística multinomial de coeficientes aleatorios; la misma que fue utilizada en el MCLM, que es

$$r_{It} = \frac{\tau_t^2}{\tau_t^2 + \pi^2/3}$$

El término $\pi^2/3$ representa la varianza del término del error que sigue una distribución logística estándar, lo cual equivale a $\pi^2/3 \approx 3.29$.

Al considerar una aproximación no paramétrica, se especifica un modelo de clases latentes para el segundo nivel. Los $T-1$ coeficientes aleatorios de las clases latentes del primer nivel son utilizados como indicadores de un modelo de clases latentes para el segundo nivel (Vermunt, 2003, 2008; Asparouhov y Muthén, 2008). Las clases latentes del segundo nivel poseen distribuciones diferentes respecto a los coeficientes aleatorios del primer nivel.

Vermunt (2003) hace énfasis en el término *no paramétrico*, indicando que esto no significa “distribución libre”. En el enfoque paramétrico se asume que los coeficientes aleatorios siguen una distribución normal; en este enfoque, el supuesto es reemplazado por el de una distribución multinomial (Vermunt, 2003), eliminando el supuesto de normalidad y reduciendo los cálculos requeridos (Muthén & Asparouhov, 2008).

En el MCLM es apropiado introducir una variable que provenga de una distribución mixta discreta en el segundo nivel, porque se puede clasificar los grupos (unidades del segundo nivel) en un número determinado de clases, como sucede con los individuos en el primer nivel, tal como lo demostraron Vermunt & Van Dijk (2001), al observar un mejor ajuste de la distribución mixta discreta en lugar del supuesto de normalidad multivariada de los coeficientes aleatorios.

Sea W_j , una variable latente discreta en el nivel 2 con $m = 1, \dots, M$ clases, donde se desconoce la pertenencia de los grupos a las clases latentes del segundo nivel. Las probabilidades condicionales de observar un patrón de respuestas específico para cada individuo dado la pertenencia del grupo

(unidad del segundo nivel) a la clase m , $P(\mathbf{X}_{ij} = \mathbf{s}/W_j = m)$ se define de la siguiente manera:

$$P(\mathbf{X}_{ij} = \mathbf{s}/W_j = m) = \sum_{t=1}^T P(Y_{ij} = t/W_j = m) P(\mathbf{X}_{ij} = \mathbf{s}/Y_{ij} = t)$$

y por el supuesto de independencia local presentado en la ecuación (3.5)

$$P(\mathbf{X}_{ij} = \mathbf{s}/W_j = m) = \sum_{t=1}^T P(Y_{ij} = t/W_j = m) \prod_{k=1}^K P(X_{ijk} = s_k/Y_{ij} = t) \quad (4.11)$$

Así mismo, puede definirse la probabilidad de un patrón de respuestas \mathbf{s} para las unidades que pertenecen a un grupo específico j , $P(\mathbb{X}_j = \mathbf{s})$:

$$P(\mathbb{X}_j = \mathbf{s}) = \sum_{m=1}^M P(W_j = m) \prod_{i=1}^{n_j} P(\mathbf{X}_{ij} = \mathbf{s}/W_j = m) \quad (4.12)$$

Reemplazando la ecuación (4.11) en la ecuación (4.12), se obtiene el modelo de clases latentes multinivel:

$$P(\mathbb{X}_j = \mathbf{s}) = \sum_{m=1}^M \left\{ P(W_j = m) \prod_{i=1}^{n_j} \left[\sum_{t=1}^T P(Y_{ij} = t/W_j = m) \prod_{k=1}^K P(X_{ijk} = s_k/Y_{ij} = t) \right] \right\} \quad (4.13)$$

Al comparar las ecuaciones (4.3) y (4.13), se observa claramente la diferencia en la estimación de un MCL multi-grupos y un MCLM desde una aproximación de efectos aleatorios no paramétrico, en el segundo caso se asume que cada grupo (unidad del segundo nivel) pertenece a una de las m clases.

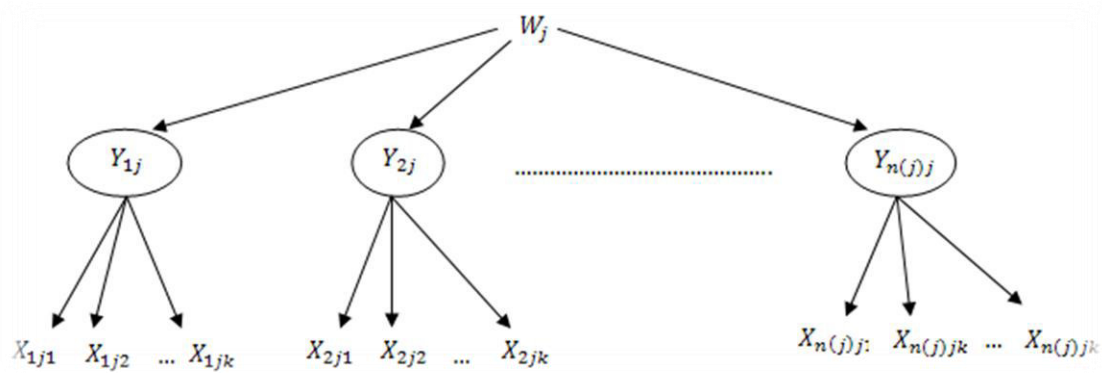


Figura 4.1. **Modelo de clases latentes con un efecto aleatorio no paramétrico en la distribución de clases latentes**

La figura 4.1 representa el MCLM, cuya variable latente del segundo nivel, W_j , influye en las n_j variables latentes del primer nivel, Y_{ij} , las cuales son mutuamente independientes dado W_j y varían por grupo; es decir, existe un Y_{ij} para cada individuo i dentro del grupo j (n_j en total) donde cada unidad del primer nivel posee su propio conjunto de indicadores, X_{ijk} .

El objetivo de la especificación del modelo es lograr una identificación significativa de las clases del primer nivel tomando en cuenta la estructura jerárquica de los datos.

La probabilidad de que el grupo j pertenezca a la clase m , se puede expresar mediante modelos log-lineales:

$$P(W_j = m) = \frac{\exp(\gamma_m)}{\sum_{r=1}^M \exp(\gamma_r)}$$

así mismo, la probabilidad de que un individuo i pertenezca a la clase latente t del primer nivel, dado la pertenencia de la clase m en el segundo nivel es

$$P(Y_{ij} = t/W_j = m) = \frac{\exp(\beta_{tm})}{\sum_{r=1}^T \exp(\beta_{trm})} \quad (4.14)$$

4.3. Estimación de Máxima Verosimilitud (EMV)

Goodman (1974) desarrolló una de las más importantes contribuciones respecto al procedimiento de estimación del MCL, donde obtuvo estimadores de máxima verosimilitud (MV) de las probabilidades de clases latentes y de las condicionales.

Los métodos numéricos más utilizados para obtener estimadores de máxima verosimilitud como se mencionó en el capítulo anterior son el algoritmo de NR y el algoritmo EM. Respecto al primer procedimiento se sabe que el método no es numéricamente muy estable y el segundo presenta dificultad para encontrar los valores iniciales para los parámetros que conducirán a la convergencia del algoritmo.

Implementación del Algoritmo EM

El algoritmo EM es un procedimiento iterativo muy eficiente que sirve para calcular los estimadores de máxima verosimilitud cuando se trabaja con un conjunto de datos incompletos.

Para un MCL estándar, la log-verosimilitud a ser maximizada es

$$\log L = \sum_{i=1}^I \log f(\mathbb{X}_i/\theta) = \sum_{i=1}^I \sum_{t=1}^T \log P(Y_i = t/\theta) f(\mathbb{X}_i/Y_i, \theta) \quad (4.15)$$

θ representa el vector de parámetros desconocidos.

Para obtener las estimaciones de los parámetros del MCLM por el método de máxima verosimilitud, Vermunt (2003) propuso utilizar el Algoritmo EM modificando el paso E.

Una implementación estándar del paso E implica el cálculo de la esperanza condicional conjunta de los n_j variables de clases latentes y las variables latentes que representan los efectos aleatorios, $P(W_j = m, Y_{ij} = t / \mathbb{X}_j)$, representando a una distribución posterior con $M^* T^{n_j}$ entradas.

En el MCLM, el valor esperado de la log-verosimilitud de los datos completos presenta la siguiente forma:

$$\begin{aligned}
 \log L &= P(\mathbb{X}_j = \mathbf{s}) \\
 &= \sum_{m=1}^M P(W_j = m / \mathbb{X}_j) \log P(W_j = m) \\
 &\quad + \sum_{m=1}^M \sum_{i=1}^{n_j} \sum_{t=1}^T P(W_j = m, Y_{ij} = t / \mathbb{X}_j) \log P(Y_{ij} = t / W_j = m) \\
 &\quad + \sum_{m=1}^M \sum_{i=1}^{n_j} \sum_{t=1}^T P(Y_{ij} = t / \mathbb{X}_j) \log P(\mathbf{X}_{ij} / Y_{ij} = t)
 \end{aligned} \tag{4.16}$$

La modificación desarrollada por Vermunt (2003) en el paso E del algoritmo EM, es utilizar las probabilidades posteriores marginales, $P(W_j = m, Y_{ij} = t / \mathbb{X}_j)$, los mismos que representan los términos que aparecen en la log-verosimilitud de los datos completos; para tal fin, calculó $P(W_j = m, \mathbf{Y}_j = \mathbf{t} / \mathbb{X}_j)$ para todo $Y_{i'j}$; $i' \neq i$, obteniendo en orden las $P(W_j = m, Y_{ij} = t / \mathbb{X}_j)$, el último término es calculado directamente.

El procedimiento se realiza bajo el supuesto de la independencia condicional, el cual se basa en el modelo gráfico asociado con la densidad definida en la ecuación (4.14), originando un algoritmo en el cual el tiempo de cálculo aumenta de forma lineal en lugar de una exponencial.

Las variables latentes son integradas empezando de las unidades del nivel inferior a las unidades del nivel superior y las probabilidades marginales posteriores se obtienen realizando el mismo procedimiento en sentido inverso.

Las probabilidades marginales posteriores $P(W_j = m, Y_{ij} = t / \mathbb{X}_j)$ son necesarias para realizar el paso E, lo que se obtiene usando la siguiente descomposición:

$$P(W_j = m, Y_{ij} = t / \mathbb{X}_j) = P(W_j = m / \mathbb{X}_j) P(Y_{ij} = t / W_j = m, \mathbb{X}_j)$$

Bajo el supuesto del MCLM, dado W_j , Y_{ij} es independiente de las variables observadas y latentes de las otras unidades del primer nivel dentro de la misma unidad del segundo nivel, lo cual resulta de que las observaciones del primer nivel son mutuamente independientes dado el valor de W_j tal como es expresado en la ecuación (4.12), se presenta la siguiente igualdad:

$$P(Y_{ij} = t / W_j = m, \mathbb{X}_j) = P(Y_{ij} = t / W_j = m, \mathbf{X}_{ij})$$

entonces $P(W_j = m, Y_{ij} = t / \mathbb{X}_j)$ se expresa de la siguiente manera:

$$P(W_j = m, Y_{ij} = t / \mathbb{X}_j) = P(W_j = m / \mathbb{X}_j) P(Y_{ij} = t / W_j = m, \mathbf{X}_{ij}) \quad (4.17)$$

donde:

$$P(W_j = m / \mathbb{X}_j) = \frac{P(W_j = m) \prod_{i=1}^{n_j} P(\mathbf{X}_{ij} / W_j = m)}{P(\mathbb{X}_j)} \quad (4.18)$$

y

$$\begin{aligned} P(Y_{ij} = t / W_j = m, \mathbf{X}_{ij}) &= \frac{P(Y_{ij} = t, \mathbf{X}_{ij} / W_j = m)}{P(\mathbf{X}_{ij} / W_j = m)} \\ &= \frac{P(Y_{ij} = t / W_j = m) \prod_{k=1}^K P(\mathbf{X}_{ij} / Y_{ij} = t)}{P(\mathbf{X}_{ij} / W_j = m)} \end{aligned}$$

La parte restante de la implementación del algoritmo EM es igual. En el paso E, la $P(W_j = m \setminus \mathbb{X})$ y $P(W_j = m, Y_{ij} = t \setminus \mathbb{X}_j)$ son utilizados para obtener mejores estimadores para las entradas de las celdas observadas en las correspondientes tablas marginales.

Para el paso M, Vermunt (2003) utilizó el método de datos completos para la estimación del modelo de regresión logística, a fin de poder actualizar los estimadores de los parámetros, usando los datos completos como si estos fueran observados.

4.4. Evaluación del Modelo de Clases Latentes Multinivel

De forma similar al MCL, las pruebas de bondad de ajuste frecuentemente empleados son el estadístico de razón de verosimilitud ji-cuadrado (L^2) y el estadístico ji-cuadrado de Pearson (χ^2). El estadístico ji-cuadrado evalúa si los estimadores de MV de las frecuencias esperadas basadas en un modelo específico difieren de las frecuencias observadas correspondientes. Cuando no se rechaza la hipótesis nula, ambos estadísticos (L^2 y χ^2) se distribuyen aproximadamente como una χ^2 ; así, el modelo presenta un buen ajuste si el valor de L^2 o χ^2 son mínimos.

El estadístico χ^2 presenta dos principales desventajas, es sensible al tamaño de muestra y surgen dificultades cuando el conjunto de datos es pequeño y el número de variables observadas es grande, obteniendo varias celdas vacías. Bajo estas condiciones, la distribución ji-cuadrada no debería ser utilizada para calcular el *p-value* porque los estadísticos de bondad de ajuste pueden desviarse considerablemente de las distribuciones conocidas (Read & Cressie, 1988).

Según McLachlan & Peel (2000), los estadísticos de bondad de ajuste aunque son utilizados para evaluar el ajuste global, no pueden ser empleados para la comparación de dos o más modelos anidados (modelos con diferentes números de clases); por lo que, propusieron el método *bootstrap*. El modelo de interés es estimado no solo por la muestra analizada sino también por n réplicas de la muestra generada de la distribución de probabilidad definida por los estimadores de MV.

4.5. Selección del Modelo de Clases Latentes Multinivel

Respecto a la selección del modelo, se prefiere que este sea simple porque provee mejores estimadores de ciertas características del verdadero modelo y tiene la ventaja de ser un modelo parsimonioso.

Los criterios de información se basan en el cálculo de índices que representan una forma penalizada de la verosimilitud, la cual se incrementa con la adición de algunos parámetros. Los criterios de información más utilizados son el Criterio de información bayesiana (BIC), Criterio de información Akaike (AIC), donde el menor valor de estos índices indica un mejor ajuste del modelo. Estos criterios de información penalizan modelos por complejidad; es decir, se puede comparar modelos con diferentes números de parámetros.

En un estudio de simulación realizado por Lukociene y Vermunt (2009) para el MCLM, mostraron que en todas las condiciones de simulación, los criterios más avanzados eran el $AIC3$ y el $BIC(k)$.

El $AIC3$, propuesto por Bozdagon (1993) se calcula de la siguiente manera:

$$AIC3 = -2LL - 3gl$$

$$LL = \sum_{i=1}^I \hat{f}(\mathbb{X}_i)$$

donde:

LL: Log-verosimilitud del modelo propuesto

gl: Grados de libertad

El $BIC(k)$, el cual representa al BIC usual, sirve para identificar la verdadera cantidad de clases latentes en los niveles superiores (Lukociene & Vermunt, 2009). A diferencia del BIC, utiliza el número de grupos (k) en la fórmula en lugar del tamaño de muestra n ,

$$BIC(k) = -2LL - gl\ln(k)$$

Al igual que los anteriores criterios de información, el menor valor obtenido por $AIC(3)$ y el $BIC(k)$ indican haber alcanzado un número óptimo de clases en el segundo nivel.

En el siguiente capítulo se presenta una aplicación del modelo de clases latentes multinivel desde el enfoque paramétrico y no paramétrico.

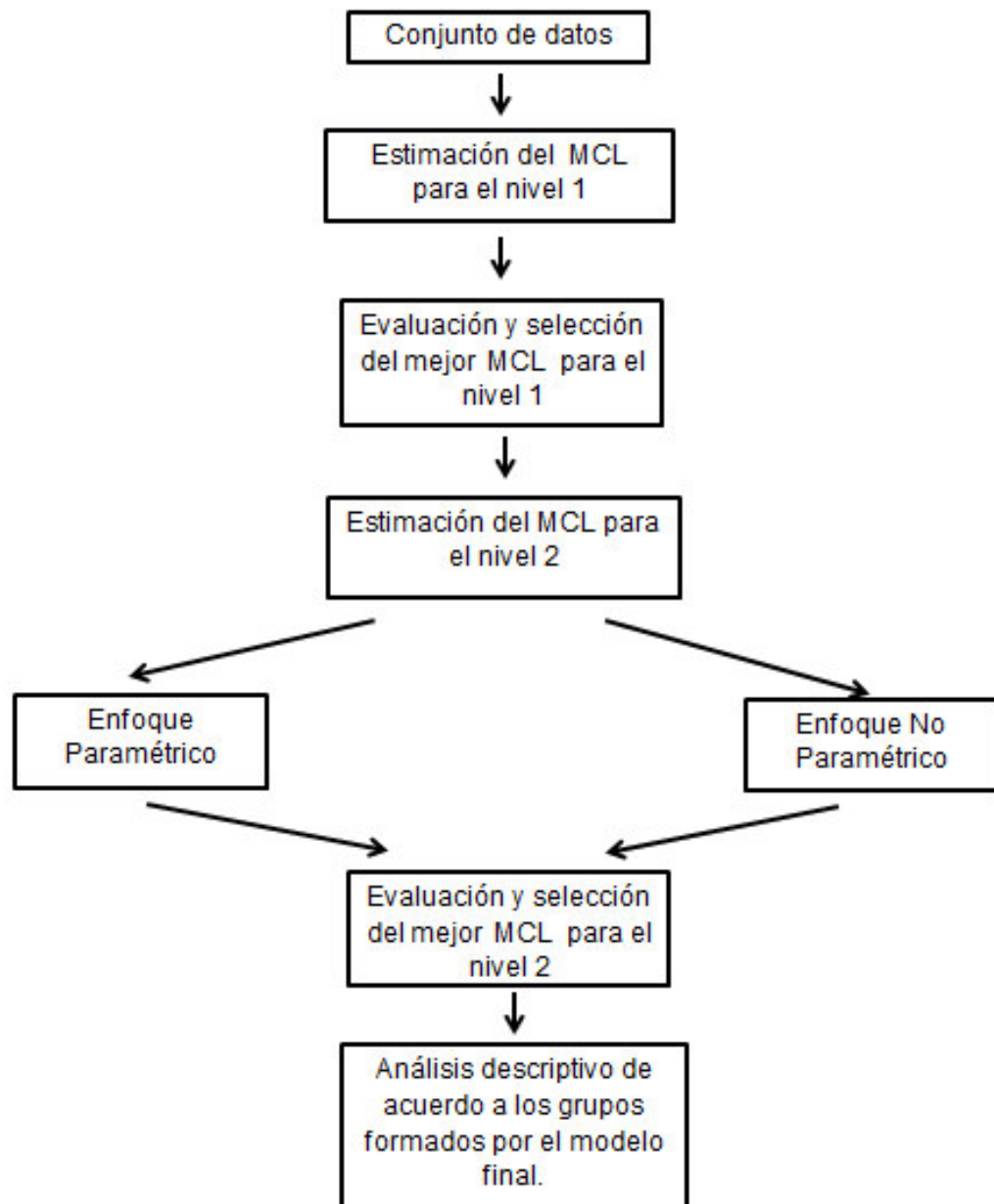


Figura 4.2. **Proceso de modelización del MCLM**

CAPÍTULO 5

Aplicación del Modelo de Clases Latentes Multinivel para la construcción de perfiles de reos pertenecientes a ocho establecimientos penitenciarios del país que sufren episodios depresivos

La depresión es uno de los mayores problemas de salud psicológica y mental que viene afectando a un gran número de personas en nuestro país, y los internos de los penales no son la excepción. Los episodios depresivos se manifiestan en diferentes niveles y se encuentra relacionado a las características de quiénes lo sufren, pero puede también estar relacionado a características de su entorno social.

Con el objetivo de tomar decisiones en la implementación de reformas que mejoren el estado emocional de los internos recluidos en los diferentes establecimientos penitenciarios, se utilizó el Modelo de clases latentes multinivel para obtener una mejor clasificación de los mismos. Para el análisis, se utilizaron los datos de internos pertenecientes a ocho cárceles del país, mostrándose en una primera etapa, la estructura de las clases latentes de los internos (unidades del primer nivel) al clasificarlos ignorando las cárceles donde se encuentran; con la información obtenida en esta etapa se realizará una segunda clasificación de los penales utilizando un análisis más complejo que modela los efectos del penal en la distribución de clases de internos que sufren episodios depresivos dentro de las cárceles.

5.1. Definición de variables

En el cuadro 5.1 se define las variables dicotómicas utilizadas según la Clasificación Internacional de Enfermedades CIE-10 para evaluar episodios depresivos, los cuales son relevantes para la formación de las clases latentes en el primer nivel.

Cuadro 5.1. Variables relacionados a episodios depresivos según el CIE-10

Variable	Descripción	Si	No
		%	%
x1	¿Le faltaba la confianza en usted mismo(a), o se sentía sin valor, inferior a los demás?	67.9	32.1
x2	¿Se hacía reproches, o se sentía culpable?	71.7	28.3
x3	¿Tenía varias veces ideas negativas como pensar que sería mejor que estuviese muerto(a), o pensaba en hacerse daño?	55.6	44.4
x4	¿Se sentía especialmente irritable (o rabioso) o se molestaba fácilmente la mayor parte del tiempo?	65.2	34.8

El Cuadro 5.2 muestra las variables que caracterizan a los internos de estos establecimientos penitenciarios, lo cual permitirá obtener una mejor visión del análisis de los datos al intervenir después de la segmentación.

Cuadro 5.2. Variables relacionadas a características de los internos

Nombre de variable	Descripción
Edad	Edad del interno
Sexo	Sexo del interno
Grado_inst	Grado de instrucción del interno
Est_civil	Estado civil del interno
Reingreso	Estado de reingreso al penal
Delito	Delito cometido por el interno

5.2. Análisis descriptivo preliminar

De un total de 399 internos de los ocho establecimientos penitenciarios; se observó que, en promedio tenían aproximadamente 35 años y la mayoría de ellos eran hombres (90%). Respecto al grado de instrucción, el 33,6% no había concluido la secundaria y sólo el 13% de ellos había logrado estudiar una carrera técnica o universitaria.

El mayor número de internos declaró encontrarse en estado de convivencia o soltería (79%); así mismo, el 30,1% de ellos se encontraba en uno de estos establecimientos más de una vez y los delitos más comunes realizados por los internos fueron contra la propiedad (41,1%) y contra la salud pública (28,6%).

Cuadro 5.3. Estadísticos descriptivos de la edad de los internos

N	399
Media	35.18
Mediana	33
Moda	25
Desviación	11.280
Asimetría	.770
Curtosis	.160
Mínimo	18
Máximo	75

Cuadro 5.4. Estadísticos acerca de las características de los internos

Variable	N	%
Sexo		
Masculino	359	90.0
Femenino	40	10.0
Grado_inst		
Sin instrucción	4	1.0
Primaria completa	35	9.1
Primaria incompleta	63	15.5
Secundaria completa	111	27.8
Secundaria incompleta	134	33.6
Superior técnico	24	6.0
Superior universitario	28	7.0
Est_civil		
Casado	49	12.3
Soltero	156	39.1
Divorciado	5	1.3
Viudo	8	2.0
Conviviente	159	39.9
Separado	22	5.5
Reingreso		
Si	120	30.1
No	279	69.9
Delito		
Vida	50	12.5
Libertad sexual	52	13.0
Libertad	4	1.0
Propiedad	164	41.1
Salud pública	114	28.6
Orden público	3	0.8
Seguridad nacional	1	0.3
Administración justicia	3	0.8
Internacionales	4	1.0

5.3. Modelo de clases latentes para los internos de ocho establecimientos penitenciarios del Perú

Se seleccionó un modelo de clases latentes estándar a partir de las cuatro variables observadas que fueron presentadas en el Cuadro 5.1, ignorándose la pertenencia del interno al establecimiento penitenciario donde se encontraba. El Cuadro 5.5 muestra los resultados del ajuste del modelo conteniendo de uno a cuatro clases latentes. Para la comparación del modelo se utiliza el valor de la log-verosimilitud (LL) y el AIC. El modelo con una clase latente es equivalente a una prueba de independencia entre los ítems, por lo cual es usado como línea de base para comparar los modelos con dos, tres y cuatro clases. El valor de la LL de los modelos varía de acuerdo a la cantidad de clases latentes que el modelo presenta, así se observa que la reducción de la LL del modelo de una clase respecto al modelo de dos clases es considerablemente significativo y la reducción del valor de la LL del modelo de dos clases respecto al modelo de tres clases es de 11.04 a 3.81. Así mismo, se observa un incremento en el valor del AIC al incorporar una clase al modelo de tres clases latentes.

Cuadro 5.5. Estadísticos de ajuste para modelos de uno a cuatro de clases latentes

Modelo	N° clases	LL	N° parámetros	AIC
1	1	106.17	4	84.17
2	2	11.04	9	-0.96
3	3	3.81	14	1.81
4	4	0.65	19	8.65

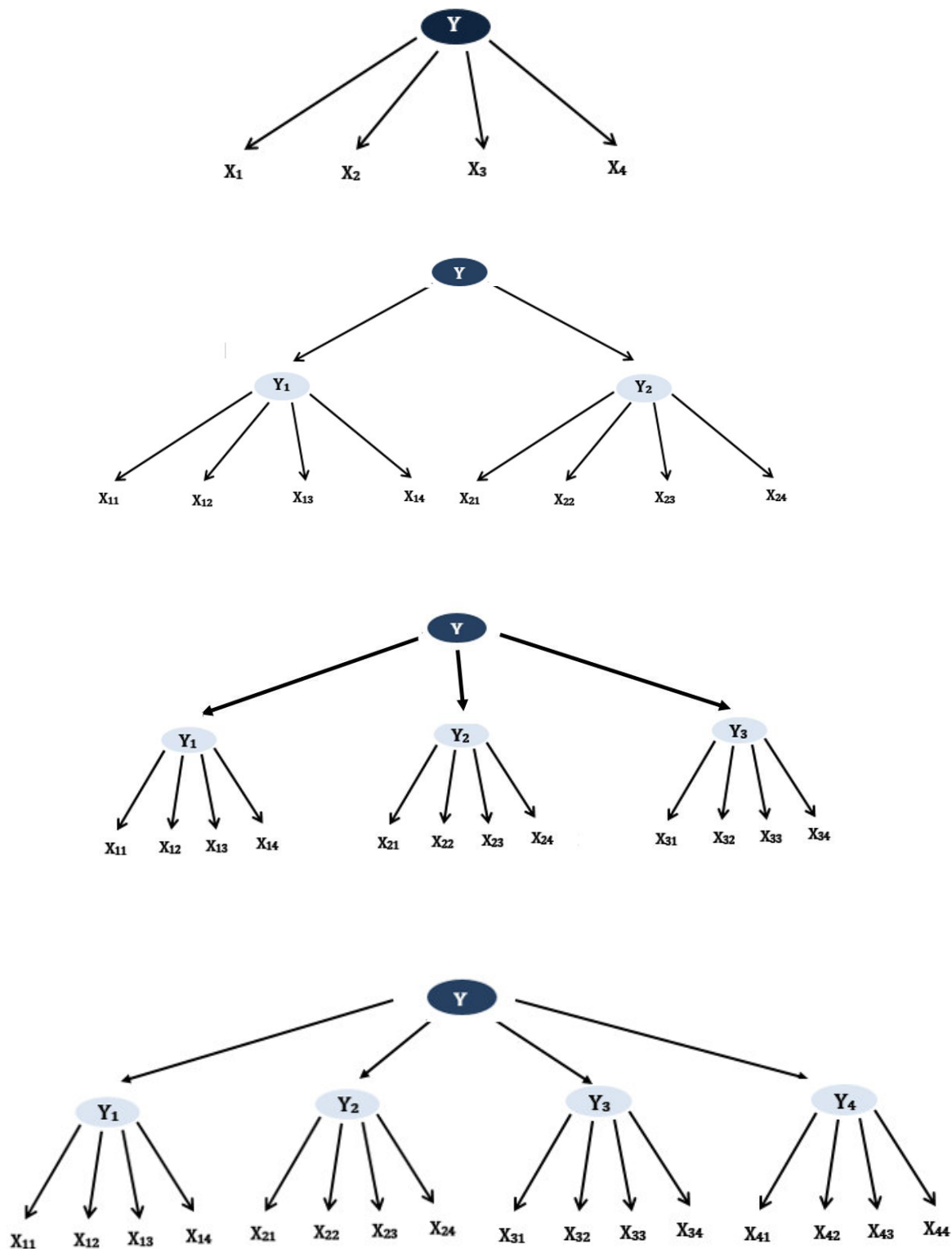


Figura 5.1. Modelos de Clases Latentes conteniendo de uno a cuatro clases latentes

Para observar las implicancias de la elección del modelo se compara los resultados de los modelos de tres y cuatro clases latentes. Los gráficos de las probabilidades condicionales para cada modelo son mostradas en las figuras 5.2 y 5.3.

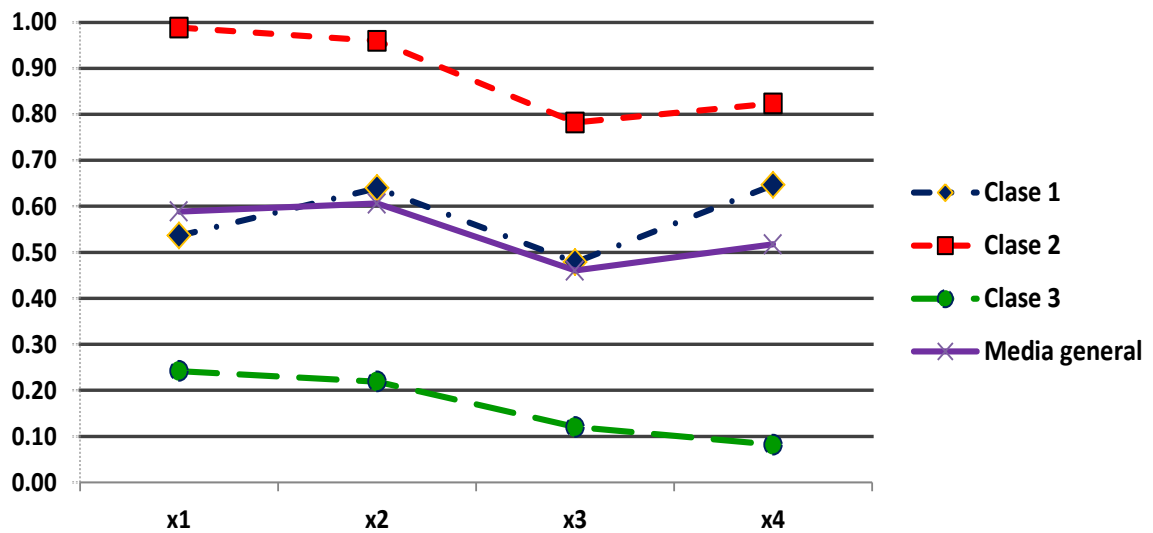


Figura 5.2. Gráfico de probabilidades condicionales de respuesta “Sí” a cada ítem, dada la clase latente para el modelo no restringido de tres clases

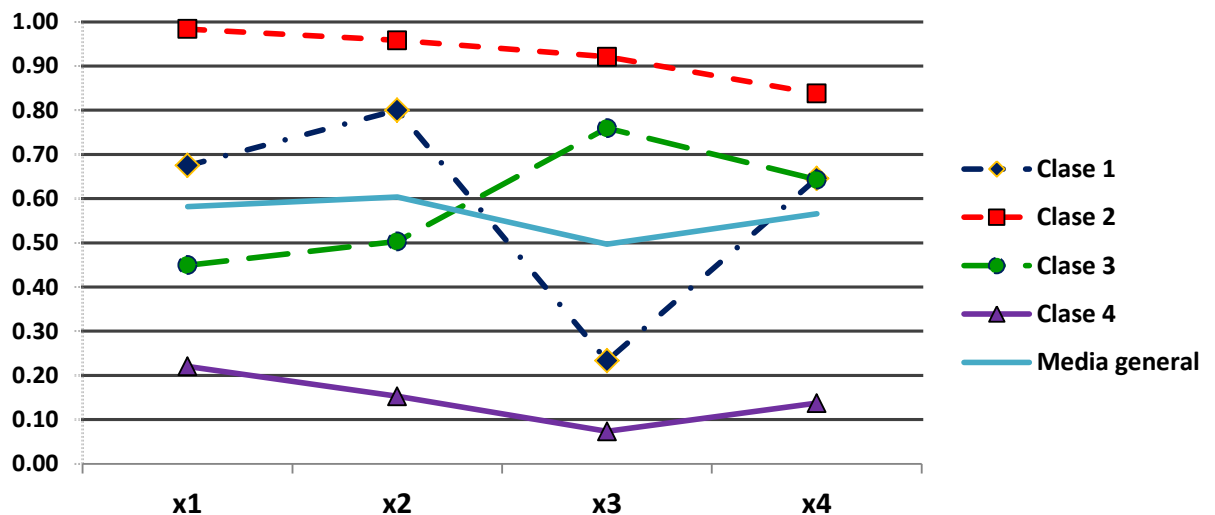


Figura 5.3. Gráfico de probabilidades condicionales de respuesta “Sí” a cada ítem, dada la clase latente para el modelo no restringido de cuatro clases

Estos gráficos muestran la probabilidad condicional de responder “Sí” a cada uno de los cuatro ítems, cada línea representa una clase latente. El modelo de tres clases mostrado en la Figura 5.2 muestra la distribución de la siguiente manera: la Clase 1, la más grande (aproximadamente 50% de la muestra total de internos) está representado por internos que tienen una moderada probabilidad de respuesta “Sí” (0.25 a 0.55) a los ítems “Le faltaba confianza en usted mismo” (x1) y “Tenía varias veces ideas negativas” (x3); la Clase 2, representada aproximadamente por el 39% de encuestados, presentaron altos niveles de episodios depresivos; mientras que la Clase 3, la cual contiene al 11% de ellos está representado por internos que raras veces o nunca experimentaron alguno de los cuatro problemas causados por la depresión. Así, se podría describir a los internos como “moderadamente afectados”, “altamente afectados” y “no afectados” por la depresión.

El modelo de cuatro clases latentes presenta algunas características similares al modelo de tres clases; las clases 2 representan lo mismo en ambos modelos. En el modelo de cuatro clases, las clases 1 y 3, representan alrededor del 54% de la muestra de internos moderadamente afectados por la depresión, lo cual es muy parecido a la clase 1 representada en el modelo de tres clases latentes. Por lo tanto, con el objetivo de buscar un modelo que represente mejor al conjunto de datos se eligió al modelo de tres clases latentes, un modelo más parsimonioso, este a su vez se utilizará para el resto del análisis de clases latentes.

5.4. Modelo de Clases Latentes Multinivel para examinar la variación a través de los establecimientos penitenciarios

Ahora se analizará la posible variación de los perfiles formados en los internos al pertenecer a uno de estos ocho establecimientos penitenciarios. Por lo que se realiza un análisis descriptivo previo de las variables relacionadas a los efectos depresivos en los internos respecto a cada penal donde se encuentran reclusos.

En el Cuadro 5.6 se muestra la cantidad de internos que integran cada establecimiento penitenciario, donde la mayoría de ellos pertenecen al penal de Lurigancho (52,6%) y sólo el 2.8%, al penal de Huaraz. Respecto a las variables relacionadas a episodios depresivos en los diferentes penales, el cuadro refleja las altas proporciones de haber sufrido al menos uno de estos episodios depresivos.

Cuadro 5.6. Estadísticos descriptivos según establecimiento penitenciario

Establecimiento penitenciario	N	%	Proporción muestral de responder "Sí"			
			x1	x2	x3	x4
Lurigancho	210	52.6	65.2	68.6	52.9	62.4
San Jorge	20	5.0	90.0	85.0	55.0	85.0
Castro Castro	20	5.0	60.0	70.0	60.0	60.0
Huaral	18	4.5	44.4	61.1	66.7	44.4
Callao	52	13.0	73.1	86.5	55.8	76.9
Huaraz	11	2.8	72.7	45.5	36.4	81.8
Ica	28	7.0	64.3	60.7	46.4	78.6
Chorrillos	40	10.0	80.0	82.5	75.0	52.5

Para la elección del modelo de clases latentes en el nivel 2, se asume que las probabilidades condicionales serán lo mismo en cada lugar; es decir, la clase latente presentará la misma estructura en cada penal.

Cuadro 5.7. Ajuste de modelos para la variación entre los centros penitenciarios (unidades del segundo nivel)

Enfoque	Modelo para los penales	N° parámetros	LL	AIC	BIC
Paramétrico	1 factor	16	-962	1956	2020
	2 factores	18	-962	1961	2032
	3 factores	20	-963	1966	2045
No paramétrico	2 clases	17	-965	1964	2032
	3 clases	20	-962	1963	2043

En un modelo de estructura latente multinivel, se presume que existe una variación entre los ocho penales; por lo tanto, ellos pueden agruparse al igual que los internos; si esto es así, debería determinarse el número de grupos.

El programa Latent Gold 5.0 permite segmentar la población en todos sus niveles, explicando la heterogeneidad entre los grupos formados. Los resultados de ajuste de cinco posibles modelos se presentan en el Cuadro 5.7. Los primeros dos modelos mostrados en la tabla plantean que los penales pueden ser clasificados dentro de dos o tres clases. El modelo de tres clases se

ajusta mejor porque el valor del AIC es ligeramente menor para este modelo respecto al modelo de dos clases. Los otros tres modelos plantean que los penales pueden estar localizados en uno, dos o tres factores continuos, de estos modelos, el de tres y dos factores no se ajustan mejor que el modelo de un factor, por lo cual se podría elegir el modelo de un factor. Como se observa en el Cuadro 5.7, los valores del AIC, BIC y LL son ligeramente menores para los modelos de factores respecto a los modelos formados por el enfoque no paramétrico. En este caso, se selecciona el modelo de tres clases formados tanto en el primer nivel (internos) como segundo nivel (penales), los mismos que difieren en términos de la proporción de sus poblaciones que están agrupados de acuerdo a los perfiles de los internos respecto al estado de depresión que padecen. Ver Figura 5.4.

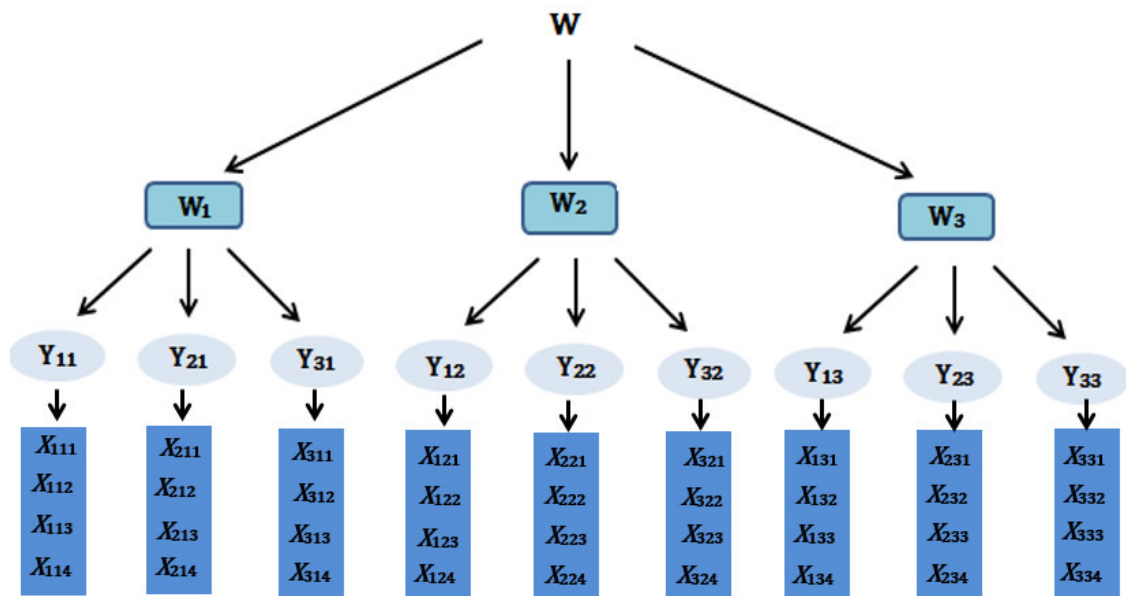


Figura 5.4. **Modelo de clases latentes multinivel con tres clases en el nivel 1 y tres clases en el nivel 2.**

El modelo de tres clases latentes para el nivel 2, agrupa a los establecimientos penitenciarios de acuerdo al mayor valor obtenido en cada una de estas clases. El Cuadro 5.8 muestra esta distribución, donde los penales de Lurigancho, San Jorge, Callao, Huaraz e Ica corresponden al primer grupo (64,9%), seguido de los penales de Castro Castro y Huaral, que conforman el según grupo (25,1%) y el penal de Chorrillos que representa al tercer grupo (10%).

Cuadro 5.8. Resultados del modelo de tres clases latentes en el nivel 2

Segmentación de penales	N	%	Probabilidades posteriores de la segmentación de penales	
			Penales	Probabilidad
1	259	64.9	Lurigancho	1.0000
			San Jorge	0.9971
			Callao	1.0000
			Huaraz	0.9461
			Ica	0.9997
2	100	25.1	Castro Castro	0.8993
			Huaral	0.9386
3	40	10.0	Chorrillos	0.9998

El modelo actual presenta sus clases latentes en dos niveles: individuos que están en una de estas tres clases y los penales, que también se encuentran en una de estas tres clases.

Con el fin de poder distinguir las clases, el término “cluster” será utilizado para el nivel individual y “Gclass” para el nivel del penal, como es manejado en el software Latent Gold 5.0.

Cuadro 5.9. Tamaño relativo de los tres clusters (clases latentes) y las probabilidades de respuesta “Sí” y “NO” para cada ítem, dado la membresía del cluster.

	Cluster1	Cluster2	Cluster3
	"Moderadamente afectados"	"Altamente afectados"	"No afectados"
Tamaño del Cluster	0.497	0.390	0.113
Variables			
Falta de confianza			
SI	0.536	0.988	0.242
NO	0.464	0.012	0.758
Sentimiento de culpabilidad			
SI	0.640	0.960	0.219
NO	0.360	0.041	0.781
Ideas negativas			
SI	0.479	0.782	0.120
NO	0.521	0.218	0.880
Sentimiento de irritabilidad			
SI	0.646	0.824	0.082
NO	0.354	0.176	0.918

El objetivo de encontrar las clases latentes (clusters) es el de poder analizar las probabilidades condicionales de pertenecer a cada una de ellas y las probabilidades condicionales de responder “Sí” a cada ítem, dado que un interno pertenezca a un cluster en particular. Las probabilidades condicionales mostradas en el Cuadro 5.9 son los valores estimados calculados considerándose el Penal al cual pertenecía el interno. La primera línea de números en la tabla muestra como los internos han sido distribuidos a través de los clusters: El primer cluster contiene alrededor del 50% de la muestra, el segundo contiene al 39% de la muestra y el último, alrededor del 11%. La finalidad de los clusters es presentar las probabilidades condicionales, las cuales son leídas bajo las tres columnas. Para cada uno de los ítems, los encuestados en el primer cluster tienen probabilidades moderadas de responder “Sí”. El segundo cluster contiene reos, quienes muestran altas probabilidades de responder “Sí”, estos internos parecen tener signos de problemas significativos debido a la depresión. El último y más pequeño cluster

tiene bajas probabilidades de responder “Sí” a cada uno de los ítems. Por lo tanto se observa que los clusters han sido distribuidos de acuerdo a internos que no han sufrido episodios depresivos, internos que han sufrido algunos episodios depresivos e internos que han sido afectados considerablemente por la depresión.

Cuadro 5.10. Tamaño de las clases del segundo nivel y distribución de encuestados entre los clusters dentro de cada clase del segundo nivel

	GClass1	GClass2	GClass3
GClass Size	0.649	0.251	0.100
Cluster1	0.583	0.310	0.150
Cluster2	0.347	0.690	0.000
Cluster3	0.069	0.000	0.850

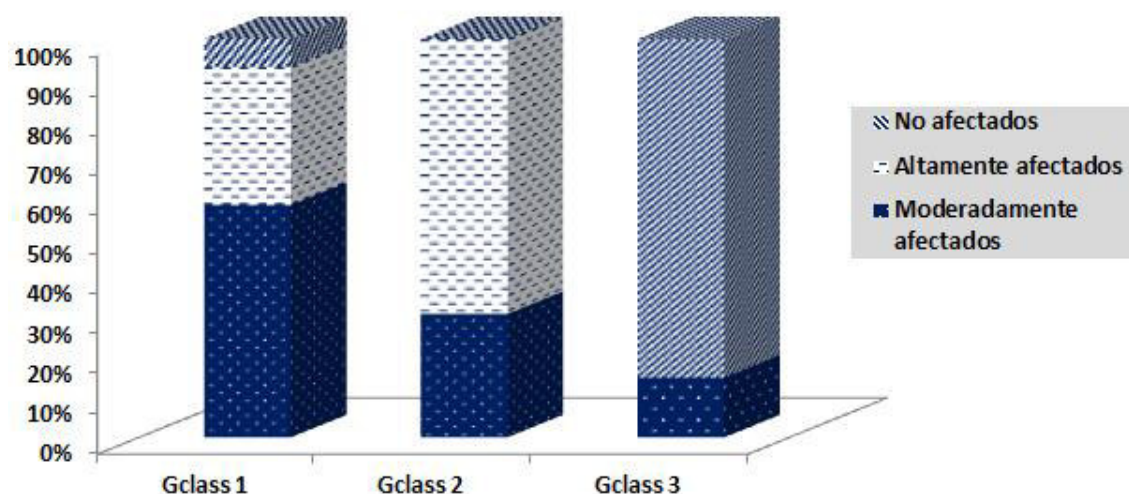


Figura 5.5. Distribución de encuestados entre los clusters latentes dentro de cada clase del segundo nivel

Los penales pueden ser agrupados dentro de las tres clases distintas, para mostrar la importancia de estas clases de los penales, la distribución de los encuestados entre los tres clusters latentes para cada uno de los grupos donde están incluidos los penales son mostrados en el Cuadro 5.10, donde las tres columnas corresponden a dichos grupos. La primera fila de números en el cuadro indica la proporción de penales que están dentro de cada grupo, así se observa que aproximadamente el 65% de los penales pertenecen al primer grupo, la cuarta parte al segundo y sólo el 10% al tercero. La parte restante del cuadro es leída en columnas, cada columna muestra la distribución de los internos perteneciendo a los penales que han sido considerados dentro del primer grupo, alrededor del 58% pertenece al primer cluster, 35% al segundo y el 7%, al tercer cluster. Los grupos que contienen a los penales (GClass) difieren totalmente en las proporciones de internos que contienen cada uno dentro de los clusters, Los penales en el GClass 1 contienen una mayor proporción de internos que han sido afectados moderadamente por episodios depresivos; mientras que los penales que pertenecen al GClass 2 presentan una mayor proporción de internos que han sido afectados de forma significativa por esta enfermedad. Los penales del GClass 3 muestran una mayor proporción de internos que no han sido afectados por estos episodios depresivos y sólo un 15% de internos que se encuentran dentro de este grupo, presenta efectos moderados de dicha enfermedad mental.

Describiendo la segmentación de los establecimientos penitenciarios (GClass)

Los centros penitenciarios que se encuentran incluidos en el primer grupo (GClass 1) está conformado en su totalidad sólo por internos, la mayoría de ellos ha alcanzado el nivel secundario; así mismo, se logra observar que un buen número de internos en estado de convivencia (43.2%) y solos (37.1%); el 65.3% de los internos de este grupo refiere ser la primera vez que ingresa a un penal; mientras que, el 47.9%, afirma haber cometido el delito contra la propiedad.

El segundo grupo (GClass 2) se caracteriza por presentar internos que al igual que el GClass 1, este grupo también está conformado sólo por internos que en su mayoría ha alcanzado el nivel secundario y viven en estado de convivencia y soltería; por otro lado, sólo el 24% de estos afirma haber ingresado a un penal más de una vez en su vida y el delito que más prevalece en este grupo es también contra la propiedad.

El pequeño y último grupo está conformado por internos del penal de Chorrillos que a diferencia de los dos grupos anteriores, todas son mujeres; el 40% de ellas ha alcanzado culminar la secundaria y un 15%, en una universidad; así mismo, se observa que la mitad de ellas refieren ser solteras y sólo un 15% de haber estado en un penal más de una vez; el delito predominante en este grupo a diferencia de los anteriores es el que cometieron contra la salud pública (87.5%). Ver Cuadro 5.11.

Cuadro 5.11. Características demográficas de los internos por grupos formados en el segundo nivel

Variable	Categoría	Gclass 1		Gclass 2		Gclass 3	
		N	%	N	%	N	%
Penal	Lurigancho	210	81.2				
	Castro Castro	20	7.7				
	Huaral	18	6.9				
	Huaraz	11	4.2				
	San Jorge			20	20.0		
	Callao			52	52.0		
	Ica			28	28.0		
	Chorrillos					40	100.0
Sexo	Masculino	259	100.0	100	100		
	Femenino					40	100.0
Grado_inst	AN (Sin educación)	2	0.8	2	2.0		
	PC (Primaria completa)	21	8.1	6	6.0	8	20.0
	PI (Primaria incompleta)	49	18.9	9	9.0	5	12.5
	SC (Secundaria completa)	66	25.5	29	29.0	16	40.0
	SI (Secundaria incompleta)	95	36.7	36	36.0	3	7.5
	ST (Superior técnico)	12	4.6	10	10.0	2	5.0
	SU (Superior universitario)	14	5.4	8	8.0	6	15.0
Est_civil	Casado	29	11.2	15	15.0	5	12.5
	Soltero	96	37.1	40	40.0	20	50.0
	Divorciado	2	0.8	2	2.0	1	2.5
	Viudo	4	1.5			4	10.0
	Conviviente	112	43.2	40	40.0	7	17.5
	Separado	16	6.2	3	3.0	3	7.5
Reingreso	Si	90	34.7	24	24.0	6	15.0
	No	169	65.3	76	76.0	34	85.0
Delito	Vida	37	14.3	12	12.0	1	2.5
	Libertad sexual	37	14.3	15	15.0		
	Libertad	4	1.5				
	Propiedad	124	47.9	39	39.0	4	10.0
	Salud pública	50	19.3	30	30.0	35	87.5
	Orden público	3	1.2				
	Seguridad nacional	1	0.4				
	Administración justicia			3	3.0		
	Internacionales	3	1.2	1	1.0		

CONCLUSIONES Y RECOMENDACIONES

- El Modelo de Clases Latentes es una metodología cuyo objetivo principal es el de segmentar la población bajo el supuesto de independencia local en t clases. El Modelo de Clases Latentes Multinivel, el cual es una extensión del MCL, tiene como finalidad seleccionar un modelo de clases latentes para cada nivel; modificando de esta manera, el supuesto de independencia presente en el MCL, característica habitual de los datos obtenidos mediante un diseño de muestras complejas. El MCLM implica una conexión directa entre ambas segmentaciones, asegurando que la cantidad de segmentos obtenidos en ambos niveles sean los más precisos y confiables para el análisis de la información.
- Los parámetros del MCLM pueden ser estimados por el método de MV. Para este propósito, se utilizó una variante del algoritmo EM propuesto por Vermunt (2003), el cual hace uso del supuesto implícito de independencia condicional del modelo, modificando el paso E del algoritmo.
- Respecto a la evaluación del MCLM se empleó los criterios de información utilizados en la selección de un MCL; tales como, el Criterio de información de Akaike (AIC) y el Criterio de información bayesiano (BIC). El criterio de información con un menor valor indica un modelo con un mejor ajuste de los datos; adicional a ellos se mencionó al AIC3 y el BIC(k), considerados como los criterios de información más avanzados según Lukociene y Vermunt (2010).
- Para los datos analizados referente a la población de internos pertenecientes a ocho establecimientos penitenciarios del Perú, se utilizó el modelo de clases latentes multinivel con la finalidad de clasificarlos de acuerdo a los episodios depresivos que estos han presentado; para tal fin se seleccionó

un Modelo de Clases Latentes Multinivel no paramétrico con tres clases latentes en el nivel 1 y tres clases latentes en el nivel 2, donde se observa que los dos niveles son modelados como interdependientes: los establecimientos penitenciarios son agrupados en base a las características similares que existen entre estos (nivel 2) a partir de los grupos formados entre los internos (nivel 1). El modelo seleccionado estratifica a los internos en tres clases o clusters, “moderadamente afectados”, “altamente afectados” y “no afectados”; a su vez, la segmentación de los establecimientos penitenciarios también se realizó en tres grupos o Gclasses, el primero conformado por los penales de Lurigancho, Castro Castro, Huaral y Huaraz; el segundo, por los penales de San Jorge, Callao e Ica, y el último sólo por el penal de Chorrillos, conformado en su totalidad por mujeres.

- Para obtener un mejor ajuste del modelo se podría incluir en el análisis, características de los individuos (covariables) en ambos niveles como predictores de las clases latentes encontradas.

REFERENCIAS BIBLIOGRÁFICAS

[1] ARAYA, C. (2010): Modelo de clases latentes en tablas poco ocupadas: Una contribución basada en bootstrap. (Tesis doctoral, Universidad de Salamanca). Recuperado de:

http://gredos.usal.es/jspui/bitstream/10366/83209/1/DES_ArayaAlp%C3%ADzar_Carlomagno_Modelos.pdf

[2] ARCHAMBEAU, C. y VERLEYSEN, M. (2004). Fully Nonparametric Probability Density Function Estimation with Finite Gaussian Mixture Models. *Université catholique de Louvain. Machine Learning Group*. Recuperado de:

http://www0.cs.ucl.ac.uk/staff/c.archambeau/publ/icapr_ca03.pdf

[3] BARTHOLOMEW, D.J. (1987), Latent Variable Models and Factor Analysis. , (2 ed) Oxford University Press, London.

[4] BAUM, C.F. (2013), Nonparametric density estimation. *EC 823: Applied Econometrics. Boston College*. Recuperado de:

<http://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn03.slides.pdf>

[5] BILMES, J.A. (1998), A Gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models. *Computer Science Division, Department of Electrical Engineering and Computer Science. U.C. Berkeley*. Recuperado de:

<http://melodi.ee.washington.edu/people/bilmes/mypapers/em.pdf>

[6] CLOGG, C. & GOODMAN, L.A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762-771. Recuperado de: <http://www.jstor.org/stable/2288706>

- [7] COLLINS, L.M., FIDLER, P.L., WUGALTER, S.E., & LONG, J.L. (1993). Goodness of fit testing for latent class models. *Multivariate Behavioral Research*, 28, 375–389
- [8] DEMPSTER, A.P., LAIRD, N.M., y RUBIN, D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1-38. Recuperado de: <http://web.mit.edu/6.435/www/Dempster77.pdf>
- [9] ESHKY, A. (2008) , Bayesian Methods of Parameter Estimation. *University of Edimburgh. School of Informatics*. Recuperado de: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/eshky.pdf
- [10] GOLDSTEIN, H. (1995). Multilevel Statistical Models. New York: Halsted Press. Recuperado de: http://www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf
- [11] GOODMAN, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61: 215-231. Recuperado de: <http://www.statmodel.com/bmuthen/ED231e/RelatedArticles/Goodman.pdf>
- [12] HABERMAN, S.J. (1979). *Analysis of Qualitative Data: New developments*, Vol. 2. Academic Press, New York. 612 pag.
- [13] HEDEKER, D. (1999). MIXNO: A Computer Program for Mixed-Effects Nominal Logistic Regression. *Journal of Statistical Software* 4:1-92. Recuperado de: http://www.jstatsoft.org/v04/i05/paper?origin=publication_detail
- [14] HEDEKER, D. (2003). A Mixed-Efeccts Multinomial Logistic Regression Model. *Division of Epidemiology & Biostatistics*. 22: 1433-1446. Recuperado de: <http://www.uic.edu/classes/bstt/bstt513/rmnom.PDF>

- [15] IZENMAN. A.J. (1991) ,Recent developments in nonparametric density estimation. *Journal American Statistics Association*, 86(413):205.224. Recuperado de: <http://yaroslavvb.com/papers/izenman-recent.pdf>
- [16] LAZARSFELD, P.F., & HENRY, N.W. (1968). *Latent Structure Analysis*. Houghton Mill, Boston. (294 pag)
- [17] LITTLE, R. y RUBIN, D. (1987), *Statistical Analysis with Missing Data. University of California at Los Angeles y Harvard University*. (1ed) John Wiley & Sons, New York. (408 pag.)
- [18] McCUTCHEON, A.C. (1987). *Latent Class Analysis*. Sage Publications, Beverly Hills. (96 pag.)
- [19] McLACHLAN, G.J. & PEEL, D. (2000) *Finite mixture models*, New York. Wiley
- [20] PERICCHI, L. R. (1987) , *Análisis de Decisión, Inferencia y Predicción Estadística Bayesiana. Centro de Estadística y Software Matemático (CESMa) y Dpto. de Cómputo Científico y Estadística*. Universidad Simón Bolívar. Caracas, Venezuela. 185pag. Recuperado de:
<http://www.pericchi.info/multimedia/documents/monografia99.pdf>
- [21] QU, Y., TAN, M., & KUTNER, M. (1996) “Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostic Tests.” *Biometrics* 52: 797-810. Recuperado de:
http://www.umgcc.org/research/models_latent_class_analysis.pdf
- [22] READ, T.R. & CRESSIE, N.A. (1988). *Goodnes of fit statistics for discrete multivariate data*, New York. Springer – Verlag, (211 pag)

[23] SEPULVEDA, R. (2004). Contribuciones al Análisis de Clases Latentes en Presencia de Dependencia Local. (Tesis doctoral, Universidad de Salamanca). Recuperado de:

<http://www.mastesis.com/tesis/contribuciones+al+analisis+de+clases+latentes+en+presencia+d:102521#sthash.LGC1iUyn.pdf>

[24] SEPÚLVEDA, R.; VICENTE-VILLARDÓN, J.L.; GALINDO, M.P. (2008). The Biplot as a Diagnostic tool of Local Dependence in Latent Class Models. A medical Application. *Statistics in Medicine*, 27(11), 1855-1869. Recuperado de:

<http://www.ncbi.nlm.nih.gov/pubmed/18265437>

[25] SINGH, A. (2005). The EM Algorithm. Recuperado de:

<http://www.cs.cmu.edu/~awm/15781/assignments/EM.pdf>

[26] SNIJDERS, T. & BOSKER, R. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Recuperado de:

<http://dx.doi.org/10.1177/0049124113494573>

[27] STUART, G y HWANG, C. R. Nonparametric Maximum Likelihood Estimation by the Method of Sieves. *The Annals of Statistics* 1982, Vol 10, No. 2: 401-414. Recuperado de:

<http://www.jstor.org/discover/10.2307/2240675?uid=3738800&uid=2&uid=4&sid=21104132050457>

[28] VERMUNT, J. & VAN DIJK, L. (2001) "A Nonparametric Random Coefficients Approach: The Latent Class Regression Model" Multilevel Modelling *Newsletter* 13: 6-13. Recuperado de:

<http://members.home.nl/jeroenvermunt/mlmnl2001.pdf>

[29] VERMUNT, J. K. (2003). Multilevel Latent Class Models. *Sociological Methodology*, 33(1), 213-239. Recuperado de:

<http://lamp2.meganet.net/~stmdlwk/technicalsupport/sm2003.pdf>

- [30] VERMUNT, J.K. & MAGIDSON, J. (2002). Latent Class Cluster Analysis. In: Hagenaars, J.A. & McCutcheon, A.L. (eds.) *Applied Latent Class Analysis*, Cambridge University Press, London, 89-106. Recuperado de: <http://www.statisticalinnovations.com/articles/lcclurev.pdf>
- [31] VERMUNT, K. y MAGIDSON, J. (2005) Technical Guide for Latent Gold 4.0. *Statistical Innovations Inc.* Recuperado de: <http://www.statisticalinnovations.com/technicalsupport/LGusersguide.pdf>
- [32] VERMUNT, L. (2008) Latent Class and Finite Mixture Models for Multilevel Data Sets. *Statistical Methods in Medical Research*. 17: 33-51. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.5854&rep=rep1&type=pdf>
- [33] VIPUL, L.B. (2010) Probability Density Estimation, *Department of Electrical Engineering*. IIT Kanpur. Recuperado de: http://home.iitk.ac.in/~mfelixor/Files/PDFestimation_NNcourse.pdf

APÉNDICE A

Modelo de Regresión Multinivel

El Modelo Multinivel fue desarrollado con el objetivo de analizar el rendimiento de los estudiantes, los investigadores llegaron a la conclusión que las observaciones de los estudiantes en el mismo salón de clases no eran independientes uno de otro; un supuesto muy común en muchas de los métodos estadísticos clásicos.

Generalmente los individuos interactúan con el grupo en el cual ellos pertenecen y a su vez, las propiedades de estos grupos están influenciados por los individuos que componen ese grupo; por lo que se les analiza como un sistema jerárquico con individuos y grupos en niveles separados. Tales sistemas pueden ser observados en niveles jerárquicos diferentes y las variables ser definidas en cada nivel.

El modelo multinivel toma en cuenta la dependencia entre las observaciones y la correlación entre los errores, supuestos que frecuentemente son vulnerados dentro de la familia de los modelos lineales generalizados, tales como el Análisis de Varianza, Correlación, Regresión y Análisis factorial.

Las variables pueden ser definidas en algún nivel de la jerarquía o ser medidas directamente en su nivel natural; así mismo, se pueden mover las variables de un nivel a otro por agregación o desagregación. Agregación significa que las variables de un nivel inferior son movidas a un nivel superior y desagregación, las variables de un nivel superior son movidas a un nivel inferior.

Incluir el modelo lineal multinivel en Modelos Mixtos Lineales puede conducir a conclusiones sustancialmente diferentes comparado con el análisis de regresión clásico.

1. Introducción

El modelo de regresión multinivel conocido también como “Modelo de Coeficientes Aleatorios” asume que existe un conjunto de datos jerárquicos, con una única variable dependiente que es medida en el nivel inferior y variables explicativas en los dos niveles. El modelo es observado también como un sistema jerárquico de ecuaciones de regresión.

El principio básico más importante es considerar que el modelo multinivel lineal puede ser visto como una extensión del modelo de regresión lineal. Sea un conjunto de n datos distribuidos en J clases grupos y n_j observaciones en cada clase, $j=1,2,\dots, J$; en el primer nivel se tiene la variable respuesta Y , y la variable explicativa X , en el segundo nivel se tiene la variable explicativa Z .

Para predecir la variable respuesta Y por la variable explicativa X , se debe obtener una ecuación de regresión por cada clase:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (\text{A.1})$$

$$E(e_{ij}) = 0 ; \text{var}(e_{ij}) = \sigma_e^2$$

Donde:

β_{0j} : Intercepto de la j –ésima ecuación de regresión.

β_{1j} : Coeficiente de regresión de la variable X de la j –ésima ecuación de regresión.

e_{ij} : Error residual de la i –ésima observación en el j –ésimo grupo.

$i: 1,2,\dots, n_j \quad j=1,2,\dots, J.$

El siguiente paso en el modelo de regresión multinivel es explicar la variación de los coeficientes de regresión β_{0j} y β_{1j} , introduciendo variables explicativas en cada clase:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (\text{A.2})$$

y

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (\text{A.3})$$

Se asume que los coeficientes de regresión γ no varían entre las clases y los errores residuales u_{0j} y u_{1j} de los modelos expresados en las ecuaciones (A.2) y (A.3) respectivamente son independientes de los errores residuales e_{ij} del modelo presentado en la ecuación (A.1); A su vez,

$$E(u_{0j}) = E(u_{1j}) = 0$$

$$\text{var}(u_{0j}) = \sigma_{u_0}^2, \quad \text{var}(u_{1j}) = \sigma_{u_1}^2, \quad \text{cov}(u_{0j}, u_{1j}) = \sigma_{u_{01}}$$

Reemplazando las ecuaciones (A.2) y (A.3) en la ecuación (A.1), se obtiene una ecuación de regresión compleja:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{1j}X_{ij} + u_{0j} + e_{ij} \quad (\text{A.4})$$

La parte determinística del modelo está formado por el segmento $[\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}]$ y la estocástica, por $[u_{1j}X_{ij} + u_{0j} + e_{ij}]$. El término Z_jX_{ij} indica la interacción entre X y Z dentro de la j -ésima clase. El efecto moderador de Z en la relación entre la variable dependiente Y y X , es expresado en una simple ecuación del modelo como una *interacción entre niveles*.

En la ecuación (4), se observa que X_{ij} y u_{1j} se multiplican, obteniendo así un error total resultante diferente por los valores de X_{ij} , lo que se conoce en análisis de regresión múltiple como *heterocedasticidad*.

Debido a que las observaciones no son independientes dentro de una misma clase, se calcula un coeficiente de correlación: la correlación intraclase (ρ).

El modelo de regresión multinivel puede ser usado para producir un estimador de la correlación intraclase. El modelo usado para este propósito no contiene variables explicativas en lo absoluto, también es llamado como *Modelo de intercepto aleatorio*; eliminando las variables explicativas de la ecuación (A.4), se obtiene la siguiente ecuación:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (\text{A.5})$$

Este modelo descompone la varianza en dos componentes independientes: σ_e^2 (varianza de los errores e_{ij} del nivel inferior y $\sigma_{u_0}^2$, varianza de los errores u_{0j} del nivel superior), usando este modelo la correlación intraclase ρ se define como:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} \quad (\text{A.6})$$

La correlación intraclase ρ indica la proporción de la varianza explicada por la estructura de agrupación en la población; la cual puede ser interpretada como la correlación esperada entre dos unidades seleccionadas aleatoriamente del mismo grupo.

Los estimadores en el análisis de regresión multinivel son estimadores de MV; los cuales estiman los parámetros del modelo maximizando la Función de Verosimilitud.

La estimación de Máxima Verosimilitud incluye procedimientos que generan errores estándar para los parámetros estimados, los que pueden ser utilizados en la prueba de significación mediante el cálculo de la estadística de prueba Z ($Z = \text{parámetro} / \text{desviación estándar del parámetro}$). Si dos modelos están anidados, significa que un modelo específico puede ser derivado de un modelo más general eliminando parámetros del modelo general, las desviaciones de los dos modelos pueden ser usados para comparar su ajuste estadísticamente. Para modelos anidados, la diferencia en las desviaciones tiene una distribución ji-cuadrada con grados de libertad igual a la diferencia en el número de parámetros que son estimados en los dos modelos.

El Modelo de intercepto aleatorio es útil como un modelo nulo que sirve como un punto de referencia con la cual otros modelos son comparados.

2. Modelo de Regresión Logística Multinivel

Al reflejar la utilidad del análisis multinivel y la importancia de las respuestas categóricas en muchas áreas de investigación, se desarrolló una generalización de los modelos multinivel. Para variables de respuesta dicotómica se diseñaron modelos de regresión probit y logística y varios métodos que incorporaron y estimaron la influencia de los efectos aleatorios.

Snijders y Bosker (1999) ofrecen un resumen del modelo de regresión logística multinivel y varios procedimientos para estimar los parámetros.

2.1. Definición del modelo

Sea j las unidades del segundo nivel (grupos), $j = 1, 2, \dots, J$; e i , las unidades en el primer nivel (las observaciones anidadas dentro de cada grupo); $i = 1, 2, \dots, n_j$. El número total de observaciones está dado por $n = \sum_{j=1}^J n_j$. Sea Y_{ij} ,

los valores de la variable respuesta dicotómica codificado con 0 ó 1, asociado con la unidad i del primer nivel dentro de la j -ésima clase.

Modelo de Regresión Logística

Sea y , una variable respuesta con distribución Bernoulli, $y \sim B(1, \pi)$ y x una observación; el modelo de regresión logística clásico desarrollado por Hosmer y Lemeshow (2000) es el siguiente:

$$y_{ij} = \pi_{ij} + e_{ij} \quad (A.6)$$

$$E(e_{ij}) = 0 ; \text{var}(e_{ij}) = \sigma_e^2 = \pi_{ij}(1 - \pi_{ij})$$

Y escrito en términos del logaritmo de los *odds ratios* de la probabilidad de respuesta, denotado por $\pi_{ij} = \Pr(Y_{ij} = 1)$ es

$$\text{logit}(\pi_{ij}) = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \alpha + x'_{ij}\beta \quad (A.7)$$

x'_{ij} es el vector $px1$ de respuestas de la i -ésima observación que pertenece al j -ésimo grupo (incluye al 1 por el intercepto), β es el vector $px1$ de parámetros de regresión desconocidos. El modelo logit como una función lineal implica que la función de probabilidad es

$$\pi_{ij} = \frac{\exp(\alpha + x'_{ij}\beta)}{1 + \exp(\alpha + x'_{ij}\beta)} \quad (A.8)$$

Se observa que los subíndices i y j reflejan el hecho de que los individuos se encuentran dentro de uno de los grupos. Sin embargo, este modelo no contiene los efectos de los grupos, no toma en cuenta la variación entre los grupos y los individuos que pertenecen a estos grupos.

Modelo de Regresión Logística Multinivel

Existen muchas maneras de extender un modelo simple a un análisis multinivel. Una manera sencilla de tomar en cuenta los efectos de los grupos es adicionar variables de diseño (variables dummy) a la ecuación (7); a su vez, se observa que cada grupo tiene su propio intercepto en el modelo. Estos interceptos son utilizados para medir la diferencia entre grupos,

$$\text{logit}(\pi_{ij}) = \alpha_j + \mathbf{x}'_{ij}\boldsymbol{\beta} \quad (\text{A.9})$$

Los interceptos pueden ser especificados como efectos fijos o efectos aleatorios (Demidenko, 2004). El uso de interceptos fijos conduce a un incremento del número de parámetros adicionales igual al número de grupos menos uno ($J-1$). Sin embargo, los interceptos, α_j ($j=1, \dots, J$) que son tratados como una variable aleatoria con una distribución de probabilidad específica, la cual conduce a un modelo de interceptos aleatorios y a estimadores más tradicionales de los efectos:

$$\alpha_j = \alpha + u_j \quad (\text{A.10})$$

En este modelo, los efectos de los grupos son medidos como interceptos aleatorios α_j ($j=1, \dots, J$), una combinación lineal de una gran media (α) y la desviación (u_j) de dicha media. Se asume que la variable aleatoria u_j sigue una distribución normal $u_j \sim N(0, \sigma_u^2)$ y la independencia entre los errores de las unidades del primer nivel (e_{ij}).

Reemplazando (A.10) en (A.9):

$$\text{logit}(\pi_{ij}) = \alpha + x'_{ij}\beta + u_j \quad (\text{A.11})$$

Se obtiene un modelo mixto porque contiene efectos fijos (α, β) y efectos aleatorios (u_j) ; este modelo es parte de la familia de los modelos mixtos lineales generalizados.

APÉNDICE B

Aplicación del Latent GOLD 5.0 para estimar el Modelo de Clases Latentes Multinivel

Para esta sección se consideró las cuatro variables descritas en el Capítulo 5 para mostrar la estimación del MCLM y las salidas más resaltantes del modelo, para lo cual se utilizó el archivo de datos “aplicación.sav”.

	ID_Cabecera	CodigoPPL	ID_Est_Penal	X1	X2	X3	X4	Region	Pabellon	Fecha_Entrevista	Edad	ID_Sexo	Fecha_Nac	Grado_Instruccion	ID_Estado Civil	Nro_Hijos	Edad_Ultimo_Hijo
1	357	102	Penal de Lurigancho	SI	SI	SI	NO	LIMA	12B	23-Dec-2008	37	M	06-Jun-1971	SI	Casado	3	14
2	358	93	Penal de Lurigancho	SI	SI	SI	NO	LIMA	12B	23-Dec-2008	50	M	24-Jun-1959	SC	Casado	4	6
3	359	91	Penal de Lurigancho	SI	SI	SI	NO	LIMA	12B	23-Dec-2008	73	M	29-Nov-1935	PI	Conviviente	2	38
4	363	92	Penal de Lurigancho	SI	SI	SI	SI	LIMA	12A	23-Dec-2008	26	M	29-Jul-1982	SI	Conviviente	1	4
5	364	90	Penal de Lurigancho	NO	SI	NO	SI	LIMA	12A	23-Dec-2008	34	M	23-Dec-1974	SI	Conviviente	4	4
6	366	71	Penal de Lurigancho	NO	SI	NO	NO	LIMA	11A	22-Dec-2008	28	M	12-May-1980	SI	Soltero	1	9
7	367	70	Penal de Lurigancho	SI	NO	SI	NO	LIMA	11A	22-Dec-2008	27	M	22-May-1981	SI	Conviviente	3	3
8	368	69	Penal de Lurigancho	NO	SI	SI	SI	LIMA	11B	22-Dec-2008	27	M	16-Jul-1981	SI	Soltero	2	6
9	369	68	Penal de Lurigancho	SI	SI	NO	SI	LIMA	11B	22-Dec-2008	26	M	10-Dec-1982	SI	Soltero	2	0
10	371	66	Penal de Lurigancho	SI	SI	NO	SI	LIMA	11B	22-Dec-2008	33	M	10-Nov-1975	PI	Soltero	1	10
11	373	75	Penal de Lurigancho	NO	SI	SI	NO	LIMA	10	22-Dec-2008	33	M	02-Feb-1975	SI	Soltero	6	1
12	378	59	Penal de Lurigancho	NO	SI	NO	SI	LIMA	05	18-Dec-2008	37	M	09-Sep-1978	SI	Conviviente	2	16
13	379	58	Penal de Lurigancho	NO	SI	NO	SI	LIMA	05	18-Dec-2008	28	M	29-Sep-1980	SI	Soltero	0	0
14	380	57	Penal de Lurigancho	SI	SI	SI	NO	LIMA	05	18-Dec-2008	46	M	04-May-1962	PC	Separado	1	18
15	382	55	Penal de Lurigancho	NO	SI	SI	NO	LIMA	05	18-Dec-2008	31	M	15-Jun-1977	SI	Conviviente	0	0
16	387	64	Penal de Lurigancho	NO	NO	SI	SI	LIMA	02	18-Dec-2008	22	M	12-Aug-1986	SI	Conviviente	0	0
17	388	63	Penal de Lurigancho	SI	SI	SI	SI	LIMA	02	18-Dec-2008	28	M	11-Sep-1980	SI	Soltero	1	3
18	389	62	Penal de Lurigancho	NO	NO	NO	NO	LIMA	02	18-Dec-2008	28	M	11-Dec-1981	PI	Soltero	0	0
19	390	61	Penal de Lurigancho	NO	NO	NO	NO	LIMA	02	18-Dec-2008	20	M	10-Apr-1988	SI	Conviviente	1	2
20	395	83	Penal de Lurigancho	SI	SI	SI	SI	LIMA	09	16-Dec-2008	36	M	04-Dec-1972	ST	Conviviente	1	11
21	398	80	Penal de Lurigancho	SI	SI	SI	SI	LIMA	09	16-Dec-2008	27	M	20-Jul-1981	SC	Conviviente	4	2
22	401	77	Penal de Lurigancho	NO	NO	SI	SI	LIMA	09	16-Dec-2008	48	M	23-Apr-1960	SU	Conviviente	8	12

Figura B.1. Archivo de datos en SPSS aplicación.sav

Abrir un archivo de datos

Abrir el archivo desde el menú en la opción File/open

Seleccionar el tipo de archivo la opción SPSS system files (".sav) y buscar el archivo de datos

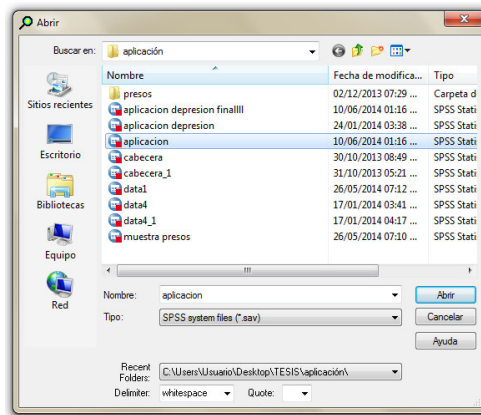


Figura B.2. Cuadro de diálogos Open File

Seleccionar aplicación.sav y retornar a la ventana principal del programa.

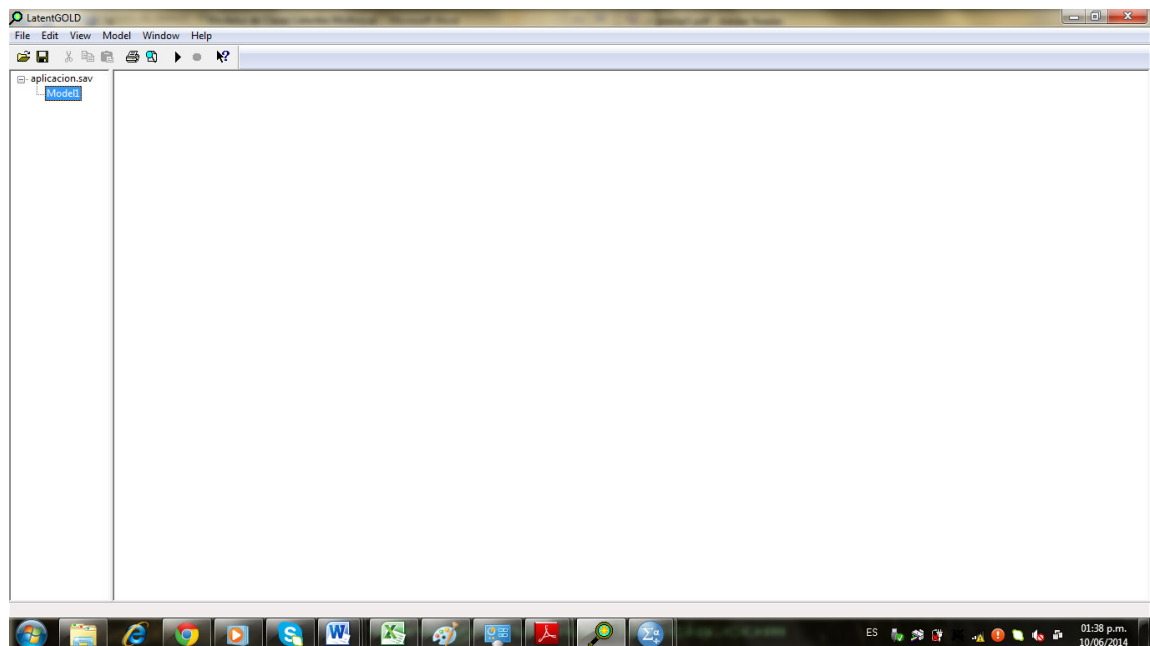


Figura B.3. Ventana principal del programa Latent GOLD 5.0

Estimación del modelo de clases latentes estándar

Dar click derecho en “Model1” para abrir el menú Model Selection

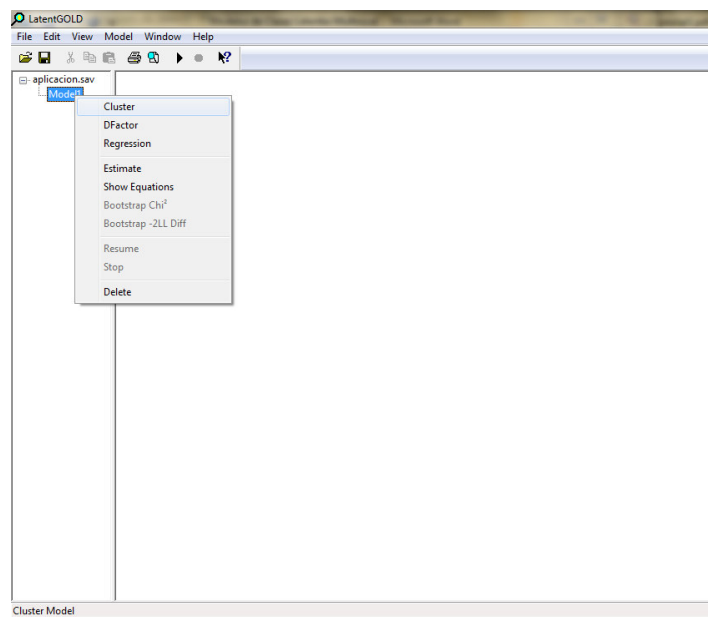


Figura B.4. Menú de selección del modelo

Seleccionar la opción Cluster y se abrirá el siguiente cuadro de diálogos para el modelo Cluster de Clases Latentes:

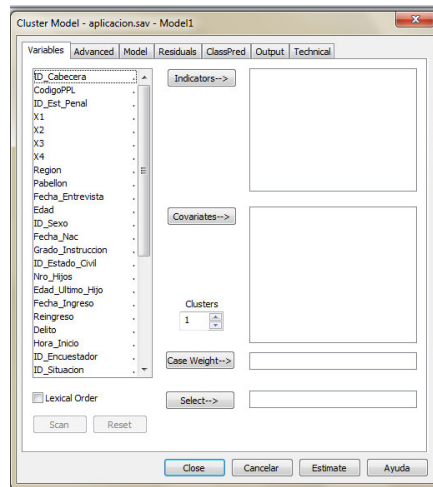


Figura B.5. Cuadro de diálogos para analizar el Modelo de Clases Latentes para el primer nivel

A continuación se ingresa las variables relevantes para el análisis en la opción “Indicators”.

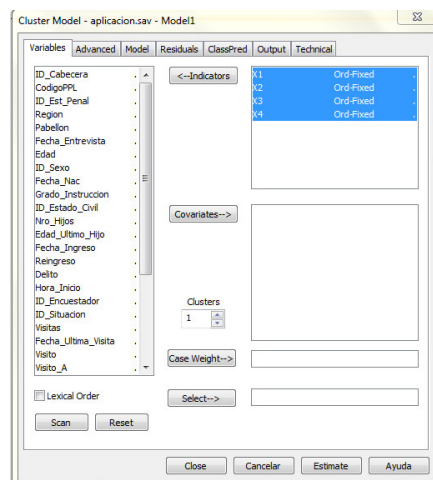


Figura B.6. Ingreso de variables

Para determinar el número de clusters se estimará cuatro modelos cluster diferentes, cada uno con un número de clusters específico. Para la aplicación se consideró los modelos que presentaban desde uno hasta cuatro clusters.

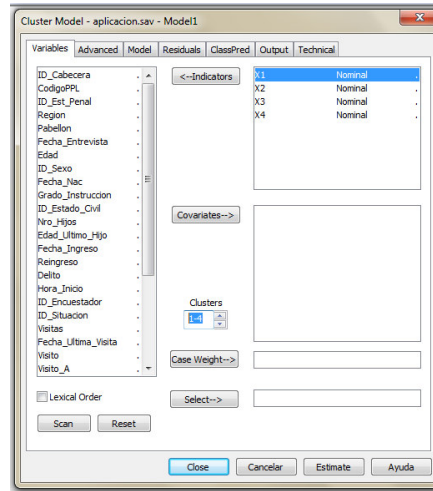


Figura B.7. **Cuadro de diálogos previo a la estimacion del modelo**

Una vez seleccionado las variables y especificado los modelos, se realiza la estimación del modelo, para lo cual se da click derecho en “Estimate”.

Cuando el programa Latent GOLD completa la estimación, el estadístico de bondad de ajuste L^2 , aparece al lado derecho del nombre asignado para cada modelo estimado.

Para poder realizar una comparación y seleccionar el modelo más óptimo se da click derecho en el nombre del archivo de datos que se encuentra en la parte superior de los modelos estimados.

Model	Cluster	LL	BIC(LL)	Npar	L ²	df	p-value	Class.Err.
Model1	1-Cluster	-1020.0979	2064.1516	4	106.1705	11	1.1e-17	0.0000
Model2	2-Cluster	-972.5332	1998.9670	9	11.0412	6	0.087	0.1473
Model3	3-Cluster	-968.9181	2021.6817	14	3.8110	1	0.061	0.2443
Model4	4-Cluster	-967.3351	2048.4604	19	0.6450	-4	.	0.2633
Model5	0-Cluster							

Figura B.8. **Resumen del modelo**

El programa Latent GOLD ofrece una opción alternativa para evaluar el modelo utilizando el *bootstrap* del L^2 .

De los cuatro modelos estimados, se seleccionó el modelo con tres clases latentes, cuya estimación de los parámetros puede ser visualizada al seleccionar la opción “Parameters”, de igual forma, se puede observar otros resultados que ayudarán a obtener una mejor interpretación del modelo.

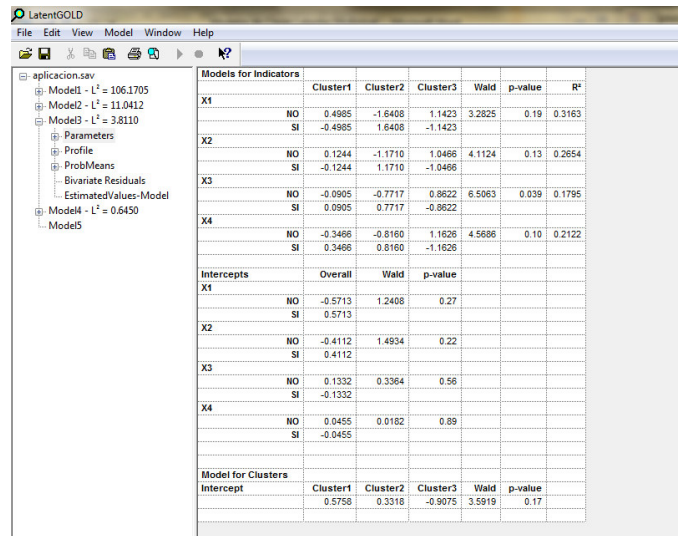


Figura B.9. Estimacion de parametros del MCL con tres clases

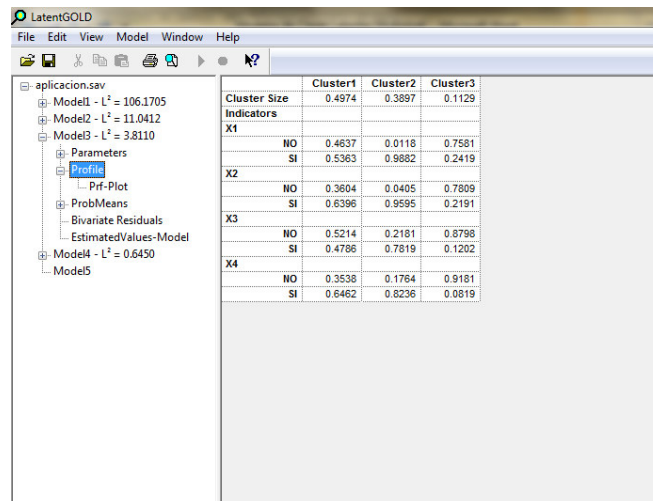


Figura B.10. Perfil obtenido en el MCL con tres clases

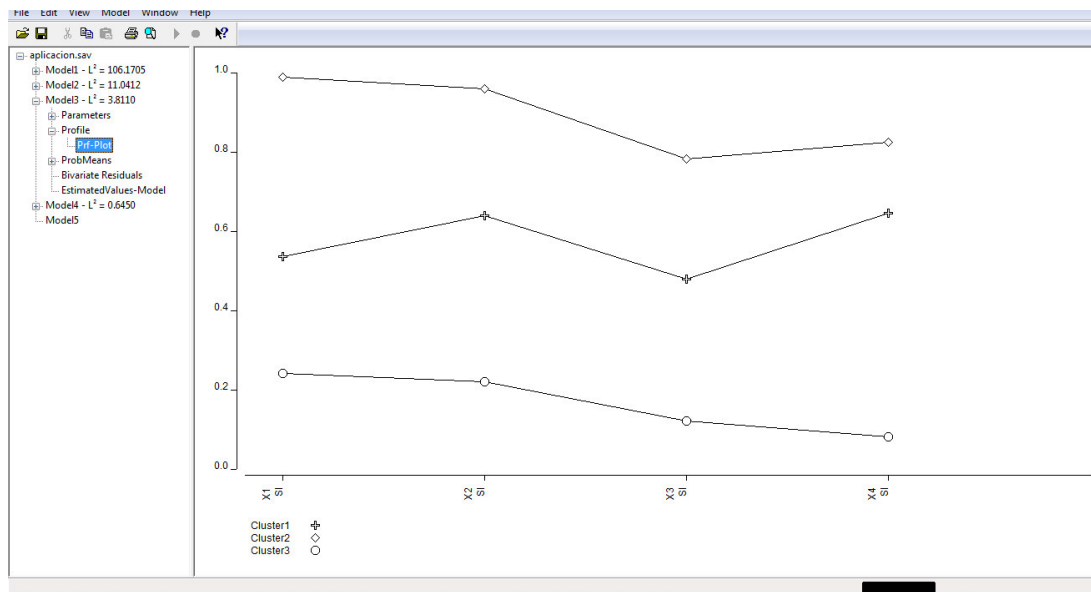


Figura B.11. Gráfico de perfiles para el MCL con tres clases

Una vez obtenido el modelo de clases latentes para el primer nivel, se selecciona el “modelo5”, el cual permite ingresar nuevamente al cuadro de diálogos para especificar un nuevo modelo, como el modelo seleccionado en el primer nivel consta de tres clases latentes, se especifica el número 3 en la opción “Clusters” y se ingresa a la opción “Advanced”, el cual se encuentra en parte superior del cuadro.

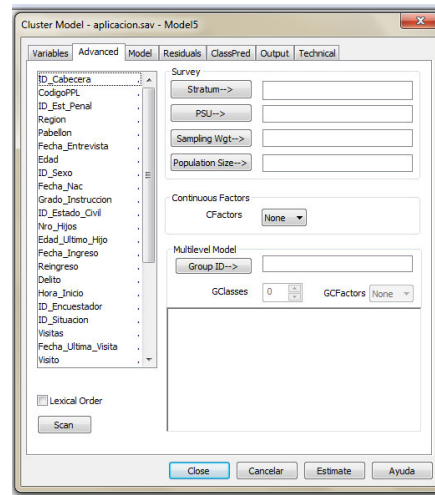


Figura B.12. Cuadro de diálogos para la estimación del MCLM

En esta ventana se especifica la variable que será utilizada para el análisis del modelo de clases latentes en el segundo nivel en la opción “Group ID” y de forma similar al procedimiento realizado para la selección del MCL en el nivel 1, en esta opción se considerarán cuatro modelos que serán especificados en la opción “GClasses”. Siguiente paso dar click en “Estimate”.

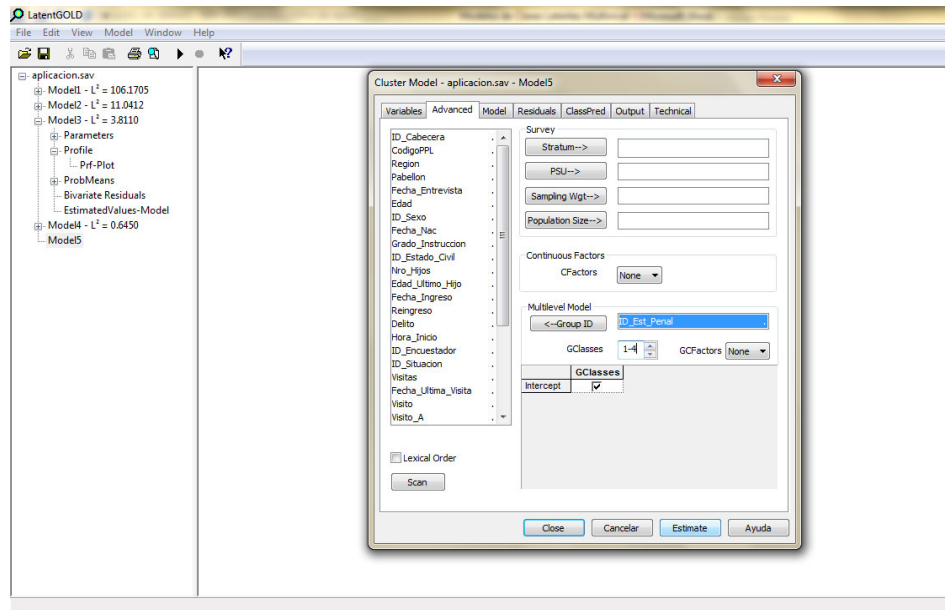


Figura B.13. Cuadro de diálogos previo a la estimación del MCLM de 1 a 4 clases en el segundo nivel

Los resultados obtenidos se presentan de forma similar al obtenido en el MCL para el nivel 1.

ANEXO

Frecuencia de datos observados

X_1	X_2	X_3	X_4	f_x
NO	NO	NO	NO	29
NO	NO	NO	SI	11
NO	NO	SI	NO	8
NO	NO	SI	SI	14
NO	SI	NO	NO	15
NO	SI	NO	SI	23
NO	SI	SI	NO	11
NO	SI	SI	SI	17
SI	NO	NO	NO	13
SI	NO	NO	SI	14
SI	NO	SI	NO	8
SI	NO	SI	SI	16
SI	SI	NO	NO	25
SI	SI	NO	SI	47
SI	SI	SI	NO	30
SI	SI	SI	SI	118

Segmentación de internos (unidades del primer nivel)

X_1	X_2	X_3	X_4	f_x	Cluster	Cluster1	Cluster2	Cluster3
NO	NO	NO	NO	29	3	0.221	0.000	0.779
NO	NO	NO	SI	11	1	0.852	0.001	0.147
NO	NO	SI	NO	8	1	0.656	0.001	0.343
NO	NO	SI	SI	14	1	0.971	0.005	0.025
NO	SI	NO	NO	15	1	0.640	0.004	0.356
NO	SI	NO	SI	23	1	0.959	0.015	0.026
NO	SI	SI	NO	11	1	0.903	0.022	0.075
NO	SI	SI	SI	17	1	0.938	0.059	0.003
SI	NO	NO	NO	13	1	0.499	0.017	0.484
SI	NO	NO	SI	14	1	0.882	0.076	0.042
SI	NO	SI	NO	8	1	0.783	0.103	0.113
SI	NO	SI	SI	16	1	0.744	0.251	0.005
SI	SI	NO	NO	25	1	0.623	0.281	0.096
SI	SI	NO	SI	47	2	0.463	0.534	0.004
SI	SI	SI	NO	30	2	0.359	0.633	0.008
SI	SI	SI	SI	118	2	0.182	0.818	0.000

Segmentación de los Centros Penitenciarios (unidades del segundo nivel)

<i>Centros penitenciarios</i>	X_1	X_2	X_3	X_4	<i>Cluster</i>	<i>GClass</i>	<i>GClass1</i>	<i>GClass2</i>	<i>GClass3</i>
<i>Lurigancho</i>	NO	NO	NO	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	NO	NO	SI	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	NO	SI	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	NO	SI	SI	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	SI	NO	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	SI	NO	SI	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	SI	SI	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	NO	SI	SI	SI	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	NO	NO	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	NO	NO	SI	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	NO	SI	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	NO	SI	SI	2	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	SI	NO	NO	1	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	SI	NO	SI	2	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	SI	SI	NO	3	3	0.000	0.000	1.000
<i>Lurigancho</i>	SI	SI	SI	SI	2	3	0.000	0.000	1.000
<i>San Jorge</i>	NO	NO	SI	SI	1	1	0.971	0.001	0.028
<i>San Jorge</i>	NO	SI	NO	SI	1	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	NO	SI	NO	1	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	NO	SI	SI	2	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	SI	NO	NO	1	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	SI	NO	SI	2	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	SI	SI	NO	2	1	0.971	0.001	0.028
<i>San Jorge</i>	SI	SI	SI	SI	2	1	0.971	0.001	0.028
<i>Castro Castro</i>	NO	NO	SI	NO	1	2	0.002	0.762	0.236
<i>Castro Castro</i>	NO	NO	SI	SI	1	2	0.002	0.762	0.236
<i>Castro Castro</i>	NO	SI	NO	NO	1	2	0.002	0.762	0.236
<i>Castro Castro</i>	NO	SI	NO	SI	1	2	0.002	0.762	0.236
<i>Castro Castro</i>	NO	SI	SI	NO	3	2	0.002	0.762	0.236
<i>Castro Castro</i>	SI	NO	NO	SI	1	2	0.002	0.762	0.236

<i>Castro Castro</i>	SI	NO	SI	NO	3	2	0.002	0.762	0.236
<i>Castro Castro</i>	SI	NO	SI	SI	3	2	0.002	0.762	0.236
<i>Castro Castro</i>	SI	SI	NO	NO	1	2	0.002	0.762	0.236
<i>Castro Castro</i>	SI	SI	SI	NO	3	2	0.002	0.762	0.236
<i>Castro Castro</i>	SI	SI	SI	SI	3	2	0.002	0.762	0.236
<i>Huaral</i>	NO	NO	NO	NO	1	2	0.001	0.781	0.218
<i>Huaral</i>	NO	NO	NO	SI	1	2	0.001	0.781	0.218
<i>Huaral</i>	NO	NO	SI	NO	1	2	0.001	0.781	0.218
<i>Huaral</i>	NO	NO	SI	SI	1	2	0.001	0.781	0.218
<i>Huaral</i>	NO	SI	NO	NO	1	2	0.001	0.781	0.218
<i>Huaral</i>	NO	SI	SI	NO	3	2	0.001	0.781	0.218
<i>Huaral</i>	NO	SI	SI	SI	3	2	0.001	0.781	0.218
<i>Huaral</i>	SI	NO	SI	NO	3	2	0.001	0.781	0.218
<i>Huaral</i>	SI	SI	NO	NO	1	2	0.001	0.781	0.218
<i>Huaral</i>	SI	SI	NO	SI	3	2	0.001	0.781	0.218
<i>Huaral</i>	SI	SI	SI	NO	3	2	0.001	0.781	0.218
<i>Huaral</i>	SI	SI	SI	SI	3	2	0.001	0.781	0.218
<i>Callao</i>	NO	NO	NO	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	NO	NO	SI	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	NO	SI	NO	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	NO	SI	NO	SI	1	1	0.999	0.000	0.001
<i>Callao</i>	NO	SI	SI	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	NO	SI	SI	SI	2	1	0.999	0.000	0.001
<i>Callao</i>	SI	NO	NO	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	SI	NO	NO	SI	1	1	0.999	0.000	0.001
<i>Callao</i>	SI	SI	NO	NO	1	1	0.999	0.000	0.001
<i>Callao</i>	SI	SI	NO	SI	2	1	0.999	0.000	0.001
<i>Callao</i>	SI	SI	SI	NO	2	1	0.999	0.000	0.001
<i>Callao</i>	SI	SI	SI	SI	2	1	0.999	0.000	0.001
<i>Huaraz</i>	NO	NO	NO	NO	1	1	0.561	0.028	0.411

<i>Huaraz</i>	NO	NO	NO	SI	1	1	0.561	0.028	0.411
<i>Huaraz</i>	NO	NO	SI	SI	1	1	0.561	0.028	0.411
<i>Huaraz</i>	SI	NO	NO	SI	1	1	0.561	0.028	0.411
<i>Huaraz</i>	SI	NO	SI	SI	2	1	0.561	0.028	0.411
<i>Huaraz</i>	SI	SI	NO	NO	1	1	0.561	0.028	0.411
<i>Huaraz</i>	SI	SI	NO	SI	2	1	0.561	0.028	0.411
<i>Huaraz</i>	SI	SI	SI	SI	2	1	0.561	0.028	0.411
<i>Ica</i>	NO	NO	NO	NO	1	1	0.893	0.000	0.107
<i>Ica</i>	NO	NO	NO	SI	1	1	0.893	0.000	0.107
<i>Ica</i>	NO	NO	SI	NO	1	1	0.893	0.000	0.107
<i>Ica</i>	NO	SI	NO	SI	1	1	0.893	0.000	0.107
<i>Ica</i>	NO	SI	SI	SI	2	1	0.893	0.000	0.107
<i>Ica</i>	SI	NO	NO	NO	1	1	0.893	0.000	0.107
<i>Ica</i>	SI	NO	SI	SI	2	1	0.893	0.000	0.107
<i>Ica</i>	SI	SI	NO	NO	1	1	0.893	0.000	0.107
<i>Ica</i>	SI	SI	NO	SI	2	1	0.893	0.000	0.107
<i>Ica</i>	SI	SI	SI	SI	2	1	0.893	0.000	0.107
<i>Chorrillos</i>	NO	NO	NO	NO	1	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	NO	NO	SI	1	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	NO	SI	SI	1	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	SI	NO	NO	1	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	SI	NO	SI	1	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	SI	SI	NO	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	NO	SI	SI	SI	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	NO	SI	NO	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	NO	SI	SI	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	SI	NO	NO	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	SI	NO	SI	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	SI	SI	NO	3	2	0.000	1.000	0.000
<i>Chorrillos</i>	SI	SI	SI	SI	3	2	0.000	1.000	0.000